

The CIERA School Change Framework: An evidence-based approach to professional development and school reading improvement

BARBARA M. TAYLOR

University of Minnesota, Minneapolis, USA

P. DAVID PEARSON

University of California, Berkeley, USA

DEBRA S. PETERSON

MICHAEL C. RODRIGUEZ

University of Minnesota, Minneapolis, USA

Introduction

School reform is at the core of today's educational policy conversation in the United States, and most of the talk is about changing primary school reading programs in order to raise student achievement. Schools are offered both rewards (e.g., monetary incentives) and sanctions (e.g., takeover by external management teams) to motivate them to improve achievement test scores by altering programs in the direction of research-based practice. This movement in the direction of research-based practice seems reasonable in light of the fact that we have a considerable body of research on what effective schools and teachers do to promote reading success in the elementary grades. We also possess a great deal of knowledge about successful school reform and the importance of professional development in that process. The missing piece for schools, however, seems to be the procedural knowledge about how to

THE PURPOSE of this naturalistic experiment was to discover the operative elements of an evidence-based reform effort that accounted for growth in student achievement. The reform was implemented in 13 schools around the United States over the course of two years. The authors studied the impact of both school-level programmatic elements and classroom-level curricular and pedagogical practices. Evidence played a role in two distinct ways: (a) in the research base that staff members were encouraged to consult to shape their local reform efforts and (b) in the data they were provided at key points concerning their progress in shaping the schoolwide reading program, altering classroom teaching practices, and improving student scores on a range of outcome measures. Through HLM analyses of school-level and classroom-level variables, the authors found that success in implementing the reform explained a small but significant proportion of the between school variance in reading growth (as measured by comprehension and fluency scores) when looking across a given year, but it explained a substantial proportion of the between school variance in reading comprehension growth when looking across a two-year period. The teachers in the high-reform-effort schools used more research-based reading instruction practices than teachers in the low-reform-effort schools and made more changes in the direction of research-based practice. Differences between the high-reform-effort and low-reform-effort schools in implementing the various components of the school change framework are discussed in light of the broader research in school reform and national and state policies regarding school change.

The CIERA School Change Framework: An evidence-based approach to professional development and school reading improvement

EL PROPÓSITO de este experimento naturalista fue descubrir los elementos operativos de una reforma basada en la evidencia que incrementó el desempeño de los estudiantes. La reforma se implementó en 13 escuelas de los EE. UU. en el curso de dos años. Los autores estudiaron el impacto tanto de elementos programáticos de las escuelas como de prácticas curriculares y pedagógicas en las aulas. La evidencia jugó un doble papel: a) en el conjunto de investigaciones que los miembros del personal podían consultar para elaborar las reformas y b) en los datos que se les proporcionaron en ciertos momentos clave para que conocieran los progresos alcanzados en el diseño de programas de lectura, la innovación de las prácticas de enseñanza y el mejoramiento de las calificaciones de los estudiantes en un conjunto de medidas. Mediante análisis HLM de variables en el nivel de la escuela y en el nivel del aula, los autores encontraron que el éxito de la implementación de la reforma explicaba una proporción pequeña pero significativa de la variancia entre escuelas en el mejoramiento en lectura (medido con pruebas de comprensión y fluidez) en el lapso de un año determinado. En cambio, cuando se consideró un período de dos años, se halló que el éxito de la reforma explicó una proporción sustancial de la variancia entre escuelas en el mejoramiento en lectura. Los docentes de las escuelas con alta dedicación a la reforma usaron más prácticas de lectura basadas en investigaciones que los docentes de las escuelas con poca dedicación a la reforma e hicieron más innovaciones incorporando prácticas con base en investigaciones. Las diferencias entre las escuelas con alta dedicación a la reforma y aquellas con poca dedicación a la reforma en la implementación de los diversos componentes de la concepción de innovación escolar se discuten a la luz de investigaciones más amplias acerca de la reforma educativa y las políticas nacionales y estatales referidas a la innovación en la escuela.

La concepción CIERA de innovación escolar: Un enfoque del desarrollo profesional y el mejoramiento de la lectura en la escuela basado en la evidencia

DER SINN dieses naturalistischen Experimentes galt dem Entdecken der operativen Elemente einer evidenzbasierten Reformbemühung, die über den Leistungszuwachs bei Schülern Rechenschaft ablegte. Die Reform wurde im Laufe von zwei Jahren in 13 Schulen quer durch die Vereinigten Staaten ausgeführt. Die Autoren studierten jeweils die Auswirkung von programmatischen Elementen bezogen auf Schulebene und lehrplanerischen und pädagogischen Praktiken auf Klassenraumbene. Evidenz spielte eine Rolle auf zwei bestimmte Weisen: (a) in der wissenschaftlichen Untersuchungsgrundlage, welche die Mitarbeiter zu konsultieren ermutigt wurden, um ihre lokalen Reformbemühungen darauf abzustimmen, und (b) in den Daten, mit denen sie bei Schlüsselpunkten ihrer Fortschritte bei der Gestaltung des schulweiten Leseprogrammes versorgt wurden, beim Verändern der Unterrichtspraktiken im Klassenraum und im Verbessern der Leistungsergebnisse der Schüler im Rahmen einer Reihe von Ergebnisbemessungen. Durch HLM-Analysen von Variablen auf Schulebene und Klassenraumbene fanden die Autoren heraus, daß der Erfolg beim Implementieren der Reform eine kleine, doch bedeutsame Proportion in der Variante der zwischenschulischen Verhältnisabweichung in der Lesesteigerung erklärte (gemessen am Bewerten von Verständnis und Flüssigkeit), wenn man dies quer durch das vorgegebene Jahr veranschaulicht; jedoch verdeutlichte sich ein kräftiges Ansteigen bei der zwischenschulischen Variante im Zuwachs beim Leseverständnis, wenn man einen Zwei-Jahreszeitraum überblickt. Die Lehrer in den hochgradig auf Reform bestrebt Schulen nutzten häufiger wissenschaftlich basierte Praktiken des Leseunterrichts, als Lehrer in den geringer auf Reform bedachten Schulen und machten mehr Änderungen in Richtung auf forschungsfundierte Praktiken. Unterschiede zwischen den hochbestrebt Reformschulen und geringen Reformbemühungsschulen in Anwendung und Ausführung der unterschiedlichen Komponenten der Schulveränderungs-Rahmenbedingung werden im Hinblick auf eine breitere wissenschaftliche Forschung in der Schulreform und der nationalen und staatlichen Richtlinien in bezug auf Schulveränderungen diskutiert.

Rahmenbedingung der CIERA Schulveränderung: Eine aussagefundierte Annäherungsmethode zur fachlichen Entwicklung und schulischen Leseverbesserung

CIERA学校変革構造：専門性開発と学校リーディング力の向上に対する証拠に基づいたアプローチ

この自然主義的実験の目的は、生徒の熟達度における伸びの原因となる証拠に基づいた改革努力の機能因子を発見することだった。改革は、2年に渡ってアメリカ合衆国の13の学校で実施された。筆者等は、学校レベルのプログラム上の要因と教室レベルのカリキュラム及び教育実践の影響を研究した。証拠は、(a)自分達の地域の改革努力を実現するために、スタッフが参考にするように促された研究基盤と、(b)学校規模のリーディングプログラムを実現したり、教室内の教育実践を変えたり、様々な結果測定で生徒の点数を伸ばしたりする際、進歩の主要地点でスタッフが与えられたデータという2つの異なる点で役割を果たした。学校レベル及び教室レベルの変数のHLM分析によって、一年間見た場合、改革の実施の成功によって、〈理解度と流暢さの点数によって測定された〉リーディング力の伸びの学校間の変数のわずかではあるが重要な部分の変数が説明されるが、2年間見た場合は、読解力の伸びの学校間の変数のかなりの部分が説明されることが判明した。改革努力の度合いが高い学校の教師達は、それが低い学校の教師達よりも、研究に基づいた指導実践を多く採用し、研究に基づいた実践の方向性をより多く変えた。学校変革構造の様々な要素を実施する際の改革努力の度合いが高い学校と低い学校間の違いが、学校改革と学校変革に関する国家及び州の方針におけるより大規模な研究に照らして議論される。

Le programme
CIERA de
réforme de
l'école : une
démarche
évaluée de
développement
professionnel et
d'amélioration de
la lecture à
l'école

CETTE EXPÉRIENCE de terrain avait pour objectif de découvrir les éléments opératoires d'un effort de réforme évalué rendant compte d'une amélioration de la réussite scolaire. La réforme a été implantée dans 13 écoles des Etats-Unis au long de deux années scolaires. Les auteurs ont étudié l'impact à la fois d'éléments de programme au niveau de l'école et de pratiques relatives aux programmes et à la pédagogie au niveau de la classe. L'évaluation a indiqué deux voies différentes : a) l'une concerne la base de recherche que les membres des équipes ont été encouragés à consulter pour mettre en forme leurs efforts locaux de réforme et b) les données apparues aux moments-clé en ce qui concerne les progrès pour mettre en forme le programme de lecture au niveau de l'école, pour modifier les pratiques d'enseignement dans les classes, et l'amélioration des résultats des élèves pour tout un ensemble de mesures d'effets. Les auteurs, au moyen d'analyses HLM de variables au niveau de l'école et au niveau de la classe, ont trouvé que le succès obtenu en implantant la réforme explique une petite mais significative partie de la variance inter-écoles des progrès en lecture (mesurés par la compréhension et la rapidité) quand on considère une année donnée, mais qu'il explique une partie importante de la variance inter-écoles dans l'amélioration de la compréhension en lecture quand on considère une période de deux ans. Les enseignants faisant un effort important de réforme scolaire ont recouru à davantage de pratiques d'enseignement de la lecture basées sur la recherche que les enseignants faisant un moindre effort de réforme et ont fait plus de changements dans la direction des pratiques basées sur la recherche. Les différences entre les écoles faisant un effort important de réforme et celles faisant un faible effort de réforme pour implanter les différentes composantes de la structure de réforme scolaire sont discutées à la lumière de la recherche plus large de la réforme de l'école et des politiques nationales et par Etats relatives aux réformes scolaires.

Базовая инновационная
схема CIERA:
профессиональное
развитие и
обучение чтению
в школе на основе
“доказательного”
подхода

Цель данного эксперимента, проведенного в естественных условиях, состояла в том, чтобы обнаружить, какие именно элементы школьной реформы, основанной на “доказательном” подходе, являются наиболее действенными и объясняют положительную динамику в успехах учащихся. Реформа осуществлялась в 13 школах в разных частях Соединенных Штатов в течение двух лет. Авторы изучили воздействие системных и управленческих элементов на уровне школы и воспитательно-методических элементов на уровне урока. “Доказательный” подход явился значимым в двух аспектах: (а) сотрудников школы всячески поощряли пользоваться исследовательской базой данных при реформировании школы на местном уровне; (б) в определенные ключевые моменты сотрудникам школы предоставлялись данные, отражавшие динамику формирования общешкольной программы развития чтения, изменение методов обучения на отдельных уроках и рост успеваемости учащихся, связанный с предпринятыми усилиями по реформированию школы. Посредством анализа переменных на уровне школ в целом и отдельных уроков в частности авторы выяснили, что у более успешных школ в плане осуществления реформ уровень роста грамотности соответственно выше (измерялось понимание прочитанного и беглость чтения), при этом в течение одного учебного года доля различий была невелика, однако в течение двухлетнего периода эта доля становится более существенной. Учителя, работающие в школах, где предпринимались значительные усилия по реформированию процесса обучения, использовали больше научно обоснованных методов обучения чтению и предприняли больше шагов в направлении научно обоснованной практики, чем их коллеги из школ, ориентированных на реформу в меньшей степени. Различия во внедрении различных компонентов базовой инновационной схемы в школах, ориентированных на реформу в большей или меньшей степени, обсуждаются в свете более широкого исследования школьной реформы, а также политики государства и отдельных штатов в плане преобразования школы.

translate this research into school and classroom practices that lead to improved reading performance for their students.

To that end, the CIERA School Change Framework was developed to help schools translate research into practice. The Internet-based framework was like other reading improvement efforts in that it directed the faculty at each of the participating schools to relevant research and professional development activities related to effective reading instruction (Allington & Johnston, 2002; National Institute of Child Health and Human Development [NICHD], 2000; Pressley, 2002; Taylor, Peterson, Pearson, & Rodriguez, 2002). The School Change Framework, however, was different from many reading improvement efforts in that it also stressed research and action steps related to the school as a unit of change. Specifically, participants were directed to two complementary but independent bodies of scholarship to assist them in building their school reform initiative: (a) research documenting the practices present in studies of effective schools—outlier schools that exceed the performance levels predicted by their demographics (e.g., Hoffman, 1991; Taylor, Pearson, Clark, & Walpole, 2000; Taylor, Pressley, & Pearson, 2002,) and (b) research examining effective approaches to school improvement in general (e.g., Fullan, 1999; Hawley, 2002). Another unique feature of the School Change Framework was that it complemented these “external” connections to research with an “internal” focus on evidence. Teachers and administrators were asked to examine not only student performance data but also data describing classroom instruction and their overall school improvement effort as they planned and implemented their reading reform (i.e., professional development, instructional practices, curriculum choices, and assessments).

Research on effective schools

Studies of high-performing, high-poverty schools have pointed to important building-level factors that must be in place in order for all children to achieve at high levels in reading. Reviewing recent large-scale studies on effective, high-poverty schools, Taylor, Pressley, et al. (2002) noted six key elements, summarized below. A major tenet of the CIERA School Change Framework is that schools striving to “beat the odds” in terms of their students’ reading achievement need to work toward goals they set related to these key elements of effective schools.

Improved student learning has been cited as an overriding priority in effective schools (Charles A.

Dana Center, 1999; Designs for Change, 1998; Langer, 2000; Lein, Johnson, & Ragland, 1997; Taylor et al., 2000). Also, schools reported a collective sense of responsibility for school improvement. Teachers, the principal, other school staff members, and parents worked collaboratively to achieve their goal of substantially improved student learning and achievement. Langer found that in effective secondary schools, teachers and administrators had a highly organized and coherent plan to increase student achievement.

Strong building leadership has been consistently cited as a key factor in effective schools (Designs for Change, 1998; Lein et al., 1997; Puma et al., 1997; Wolf, Borko, Elliott, & McIver, 2000). Principals have been noted to engage in a range of facilitative activities—redirecting people’s time and energy, developing a collective sense of responsibility for school improvement, securing resources and professional development for teachers, providing opportunities for teachers to collaborate, increasing instructional time, and helping the school staff persist in spite of difficulties. Wolf and her colleagues, for example, found that effective leaders listened, kept an open mind, and fostered shared leadership within their buildings. This current body of work is reminiscent of earlier work summarized by Hoffman (1991) stressing the importance of strong instructional leadership. In one elaborate case study of two schools, Venezky and Winfield (1979) found that the principal set high expectations for students’ reading achievement and worked closely with a reading specialist to improve the school reading program.

In addition to, or perhaps because of, strong leadership, *strong staff collaboration* has been highlighted in studies of effective schools (Charles A. Dana Center, 1999; Designs for Change, 1998; Lein et al., 1997; Taylor et al., 2000; Wolf et al., 2000). In effective schools, teachers plan and teach together, with a focus on how to best meet students’ needs. They have reported a strong sense of building communication, talking and working across, as well as within, grades, which contributed to better understanding of one another’s curricula and expectations.

Studies of effective schools have stressed *ongoing professional development* and the implementation of research-based practices (Charles A. Dana Center, 1999; Designs for Change, 1998; Langer, 2000; Lein et al., 1997; Taylor et al., 2000). Many successful schools have emphasized a type of sustained professional development in which teachers learn together within a building and collaborate to improve their instruction.

Teachers in effective schools systematically *share student assessment data*, usually on curriculum-embedded measures, as a part of the process of making instructional decisions to improve pupil performance (Charles A. Dana Center, 1999; Designs for Change, 1998; Lein et al., 1997; Taylor et al., 2000). Teachers in effective schools also work together to carefully align instruction to state or district standards and assessments.

Effective schools have reported strong efforts within schools to *reach out to parents* (Charles A. Dana Center, 1999; Designs for Change, 1998; Lein et al., 1997; Puma et al., 1997; Taylor et al., 2000). Effective schools work to build parents' trust and then develop effective partnerships with them in order to support student achievement. Parents are valued members of the school community. Finally, effective schools also report a positive school climate, good relations with the community, and high levels of parental support.

School improvement and professional development

Research on effective school improvement and teacher professional development is consistent with the research on effective schools in general. It has stressed the importance of teachers learning and changing together over an extended period of time as they reflect on practice and implement new teaching strategies (Bryk, Sebring, Kerbow, Rollow, & Easton, 1998; Fullan, 1999; Fullan & Hargreaves, 1996; Langer, 2000; Louis & Kruse, 1995). Schools that have had successful school improvement efforts typically operate as strong professional learning communities, with teachers systematically studying student assessment data, using the data to modify their instruction, and working with colleagues to refine their teaching practices (Fullan). Reflective dialogue, deprivatization of practice, and collaborative efforts all enhance shared understandings and strengthen relationships within a school (Langer; Louis & Kruse). A second major assumption of the CIERA School Change Framework is that schools aspiring to significantly improve students' reading achievement must be informed by this research on effective school improvement and professional development if they are going to succeed as a school.

To improve instruction and performance, schools must also adopt an attitude of continuous improvement (Fullan, 2002) as well as a sense of shared commitment to the process (Newmann, 2002). To help teachers transform their reading

instruction, schools are encouraged to become learning communities (Killion, 2002; Lieberman & Miller, 2002). Valli and Hawley (2002) concluded that to be effective, professional development activities must be school based, ongoing, and tied directly to teachers' efforts to implement new or revised strategies within the classroom. They suggested two additional features for maximum effectiveness: (a) Professional development should make use of data on student work, outcome measures, and teachers' instruction, and (b) it should follow a change process that helps solve problems and moves the agenda forward.

Effective reading instruction

Research on effective reading instruction is extensive and has examined both curricular (what people do) and process (how they do it) variables. A third major tenet of the CIERA School Change Framework is that teachers wishing to improve their teaching of reading must be informed by both of these bodies of research about effective reading instruction. Upon reviewing research primarily focused on curricular aspects of reading instruction, the National Reading Panel (NRP) concluded that an effective reading program included the following: direct instruction in phonemic awareness; explicit, systematic phonics instruction; guided, repeated oral reading; direct and indirect vocabulary instruction; and comprehension strategies instruction. Many other sources, such as *The Report of the National Academy of Education on Preventing Reading Difficulties in Young Children* (Snow, Burns, & Griffin, 1998) and contributors to the *Handbook of Reading Research*, Vol. 3 (Kamil, Mosenthal, Pearson, & Barr, 2000), *Handbook of Early Literacy Research* (Neuman & Dickinson, 2001), and *Handbook of Research on Teaching the English Language Arts* (Flood, Lapp, Squire, & Jensen, 2003), have corroborated and elaborated upon these findings.

The National Reading Panel report has understandably received the lion's share of attention during the current reading reform movement, and that attention has been useful in specifying the content, or the *what*, of reading instruction. However, we contend that to significantly improve students' reading achievement, teachers must also consider the broader scope of research summarizing the pedagogical practices of effective teachers of reading—the *how* of reading instruction. Summarizing research on effective teaching processes relevant to reading achievement in the 1970s (e.g., Brophy, 1973; Dunkin & Biddle, 1974; Flanders, 1970; Stallings &

Kaskowitz, 1974), Hoffman (1991) concluded that more effective teachers focused on academics, had high numbers of pupils on task, and provided direct instruction that included making learning goals clear, asking students questions as part of monitoring their understanding of what was being covered, and providing feedback to students about their academic progress.

Research on effective pedagogical reading practices has focused on the cognitive processes used by excellent teachers. Duffy et al. (1987) found that effective teachers engaged in modeling and direct explanation to teach students strategies they could use to decode words and understand texts. Langer (1999) found that in higher performing schools, teachers explicitly taught students strategies for thinking about ideas and completing activities.

Taylor et al. (2000) found that accomplished primary-grade teachers had a preferred teaching style of coaching as opposed to telling, whereas the reverse was true for less accomplished teachers. Allington and Johnston (2002) also found that exemplary fourth-grade teachers used coaching as a teaching style to lead their students into discussion and inquiry as they collectively constructed meaning in response to text.

Taylor, Pearson, Peterson, and Rodriguez (2003) found that more effective teachers engaged students in more higher level responses to text (both in discussions and written assignments) as a part of what the researchers labeled a framework of instruction promoting cognitive engagement during reading. In addition to higher level questioning, cognitive engagement involves three additional practices: (a) teaching students word recognition and comprehension strategies that the students can use on their own during reading, (b) promoting active rather than passive student response activities, and (c) coaching rather than telling as a primary interaction strategy. Taylor et al. (2003) interpreted their findings as reminiscent of the work of Knapp and associates (Knapp, 1995), who found that effective teachers of low-income children stressed higher level thinking skills in addition to lower level skills in teaching “for meaning.” In a similar vein, Allington and Johnston (2002) reported that effective fourth-grade teachers were meaning centered, and they engaged their students in authentic conversations. Langer (1999) found that teachers in higher performing high schools taught basic skills but also moved beyond the basics to engage students in activities that involved deeper understandings.

The work of Pressley and his colleagues (Pressley et al., 2001) on exemplary teachers of read-

ing has stressed the importance of both the curricular and process aspects of effective primary-grade reading instruction. They found that effective primary-grade teachers provide a balanced, motivating literacy program in which they teach skills and strategies but also actively engage their students in a great deal of actual reading and writing. In addition, these effective teachers fostered self-regulation in students’ use of strategies when reading or writing on their own.

Significance and purposes of the current project

Many studies of reading improvement have focused on the impact on students’ reading growth of either school-level or classroom-level factors. In the current study, however, we examined the impact of both school-level elements (e.g., school effectiveness characteristics, such as leadership, collaboration, and home-school links, and overall reform fidelity) and classroom-level curricular and pedagogical practices within a set of schools implementing the CIERA School Change Framework. The purposes of this study were twofold. First, we wanted to determine the effectiveness of the School Change Framework as a structure for school reform and professional development in reading; to evaluate this aim, we would investigate whether the framework, as operationalized by schools implementing the reform effort successfully, had a positive impact on students’ reading and writing growth. Second, we wanted to determine the specific classroom- or school-level factors that accounted for students’ growth in reading and writing achievement in schools that were attempting to improve their teaching of reading.

Method

Participants

A total of 13 schools (located in Connecticut, North Carolina, Iowa, Minnesota, and California) were studied, with 8 schools in their second year in the project and 5 in their first year. Schools were high-poverty sites with a mean of 81% (and range of 70–95%) of their students qualifying for subsidized lunch. Seven of the schools were in large urban areas, three were in towns of fewer than 100,000 people, and three were in rural areas. Across the 13 schools, 20% of students (with a range of 0–78%) were English-language learners (ELLs), and 71% of

students (with a range of 48–92%) were students of color.

At least 75% of the K–5 teachers per school had agreed by secret ballot to participate in the project. Two teachers per grade were randomly selected for classroom observations and interviews. Within these designated classrooms, teachers were asked to divide their classes into thirds (high, average, and low) in terms of perceived reading ability. Nine children were randomly selected as students to be assessed, three from each band of perceived reading achievement. To study the impact of the reform effort on students' literacy growth, we analyzed data on children who had taken the same tests in the fall as in the spring; this required us to eliminate students from kindergarten and grade 1. Thus, this article focuses on students in grades 2–5 who were assessed in fluency, reading comprehension, and writing performance in October and May. A total of 92 teachers and 733 students were included in the statistical analyses.

Student assessments

We used three complementary measures of written language growth, each emphasizing a different aspect of this complex phenomenon—a standardized reading comprehension test, an index of fluency (words correct per minute [WCPM]) on a common passage, and a directed writing assessment. (In an independent effort in a single school in the overall study [Cervetti, Jaynes, & Pearson, 2002], we examined the relationship between wcpm scores and scores derived from the application of the NAEP fluency rubric. Since the correlation between these independent indexes of fluency was so high, we felt justified in using the simpler index of fluency throughout our work.) The comprehension and fluency measures gave us indexes of two faces of reading, meaning-construction processes and word-level processes, and the writing measure allowed us to determine whether the school- and classroom-level practices we observed transferred to an independent but related measure of written language competence. These general measures served our purposes well. Because the focus of the reform at any given school was up to the staff at that school, it was impossible to specify curriculum-embedded measures that would serve all of our sites. Hence, we had a preference for more general measures—but nonetheless measures on which we would expect to see growth in response to good instruction.

In the fall, all students were individually assessed in reading fluency. They read aloud for one

minute to obtain a score for the number of words read correctly in one minute (Deno, 1985), reading a passage that was one grade level below their grade placement from the 1997 Basic Reading Inventory (BRI). In a group setting students completed the reading comprehension subtest of the Gates–MacGinitie Reading Test. They also responded to a writing prompt (1998 Michigan Literacy Progress Profile). In the spring, these three assessments were repeated; however, the fluency assessment was taken while reading an on-grade-level passage (1997 BRI). We had used a below-grade-level passage in the fall to minimize the number of students reading at a frustration level. It should be noted that even with a change in passage difficulty, fall and spring scores were highly related (Pearson r of .82, .79, .78, and .80 at grades 2–5, respectively). Fall and spring passages used at each grade level were from the same form of the BRI, of the same length, and similar in content and genre.

Each response to the writing prompt was scored according to a 4-point rubric by one person from a team of trained scorers. Twenty-five percent of the writing samples at each grade level were scored by a second scorer, with 83% agreement between the two scorers.

Implementing school reading reform activities

The reading reform activities focused on professional development and school improvement. We asked that each school form a school leadership team, made up of teachers, the principal, and an external facilitator (who spent a minimum of eight hours a week in the school). The role of the leadership team was to coordinate the large- and small-group reading reform activities based upon the two data sources (the national research base and an annual school report—described below—provided by the research team). Staff members were expected to meet as a large group at least once a month for an hour and meet in hour-long study groups three times a month.

Recommended large-group activities focused on school improvement and included discussion and action on issues related to shared leadership, the schoolwide reading program, and parent partnerships. Also, schools were encouraged to have reports from study groups at the large-group meeting to foster communication and cross-grade dialogue.

Small-group activities focused on professional development in the teaching of reading. Teachers

were expected to meet regularly in within-grade and across-grade study groups to improve reading instruction. Study groups had agreed to examine aspects of classroom reading instruction supported by research (e.g., comprehension strategies instruction, phonemic awareness instruction) and implicated in the school-level report. First, study groups had committed to developing action plans specifying what they were focusing on, what would be done, when activities would be completed, and how successes of the study group would be measured. Then, study-group members across multiple sessions were encouraged to engage in the following activities: discussing research-based articles on effective practices to teach reading, watching and discussing video clips of effective practice, sharing videotapes of their own practice, problem solving, and sharing expertise related to these activities. Groups were encouraged to review information on the CIERA School Change website designed for the project. The website contained research summaries on effective reading instruction, effective schools, and effective school improvement as well as downloadable articles for teachers to discuss on research-based reading practices related to their study group's focus area. The website also contained video clips of effective practice and suggested study-group activities. To illustrate, one study group focusing on comprehension strategies read an article from *The Reading Teacher* on teaching reading strategies to third-grade students (Bergman, 1992) and then watched several two-minute video clips of a teacher teaching students how to summarize sections of an informational text on germs and to ask for clarification about things in the text that confused them.

Documenting school characteristics, reform effort, and classroom practices

School characteristics

Teachers were interviewed (about 30 minutes per interview) in the fall, winter, and spring; principals were interviewed in the fall and spring. The sets of interview data were used to document reading program features and participant beliefs. To evaluate the degree to which factors previously found to be important in effective schools existed in a specific school, we applied a five-factor coding rubric to each set of interviews. The factors were (a) building collaboration in the delivery of reading instruction; (b) links to parents; (c) reflection and change pertaining to reading instruction; (d) collaborative professional

development; and (e) building strong leadership (and the extent to which this leadership was invested in the teachers, as well as the principal). The 4-point rubric was designed to capture the strength of evidence (from our interviews) suggesting that each factor was present in that school (0 = very low or nonexistent presence, 1 = low, 2 = moderate, and 3 = high). This coding rubric is presented in Table 1. All of the sets of interviews were coded by one member of the research team. A second team member independently coded the interviews from a random sample of 25% of the teachers; the mean agreement on overall rubric scores was 87% across the five factors. The five ratings were summed to generate a school effectiveness score for each school in the study. The range of summed scores varied from a low of 4.3 to a high of 10.0, out of a possible score of 15 ($M = 7.42$, $SD = 1.96$).

School reform effort

Teachers meeting in study groups were asked to develop an action plan and to complete a common form for study-group meeting notes after each session. The external facilitators were asked to keep brief monthly logs summarizing the activities pertaining to the school change project that had transpired at their school. They were also asked to write an end-of-year report. The data from the notes, action plans, logs, and end-of-year report were used to document the reform effort at the school level.

Although schools had agreed, in principle, to the conditions of the study, they exhibited considerable variability in the degree to which they adhered to the School Change Framework. Actions important to the framework included the following: (a) meeting for one hour three times a month in study groups; (b) meeting in cross-grade study groups; (c) reflecting on teaching in study groups; (d) considering research-based "best practices" in study groups; (e) completing and being guided by action plans in study groups; (f) selecting substantive topics for study and maintaining topics over time; (g) meeting as a whole faculty once a month to set goals based on data (e.g., school report data, student reading and writing data) and to share study-group activities; (h) working on parent partnerships; (i) making effective use of the external facilitator; and (j) having an effective internal leadership team. Using the comments of each teacher across the three interviews, the study-group meeting notes, study-group action plans, facilitator logs, and the end-of-year reports, we built a scale indicating whether or not a school was perceived to be implementing the various components

TABLE 1
RUBRIC FOR RATING INTERVIEW RESPONSES

0 _____ 1 _____ 2 _____ 3
Low _____ High
A. Building collaboration (perception):
0—Teachers work in isolation or talk only at grade level, some sense of negative climate.
1—Only or mostly grade-level talk, ambivalent climate, nothing mentioned about collaboration or a learning community, or it is mentioned only in passing.
2—Some talk across grades but not a great deal, collaboration is mentioned but not stressed, teachers provide specific examples of how they are collaborating within their building, some sense of positive climate.
3—Cross-grade talk, collaboration on delivery of reading program, on professional development, collaborative learning community, positive climate.
B. Links to parents (school's efforts to reach out to parents):
0—Teachers expressed considerable dissatisfaction with parental involvement and little or nothing is being done by the school to facilitate a link with students' home environments.
1—Very little mentioned about parents, or teachers expressed dissatisfaction with parental involvement.
2—Some teachers actively pursue parental involvement in the classroom, mention that parents participate in opportunities offered at school (i.e., library reading program, parent center, site council, school meetings).
3—Includes those activities listed in B2 rating, but also includes a schoolwide focus, with teachers conducting phone or written surveys, interviews or focus groups to find out parents' concerns; teachers or principal calling home at least once a month with good news as well as to discuss concerns; teachers sending home a newsletter or personal note at least once a week; anything else that the school does to invite parents in as partners.
C. Instructional reflection and change:
0—Little or no reflection on instructional practice by the individual classroom teachers, some talk between individual teachers about what is working.
1—Teachers talk and share ideas with one another about what is working in their classrooms during formal meeting times (e.g., grade-level meetings).
2—Teachers talk and share ideas with one another in study groups. They may examine student work, reflect on their own instructional practice, and read current research on best practices, but most of their discussions focus on sharing what they do in their own classrooms.
3—Teachers indicate they are intentionally reflecting on their practice and are seriously working with others to improve their practice (i.e., study groups with action plans, grade-level meetings to improve instruction), discussion within groups is informed by research on best practices and student assessment data.
D. Views of professional development:
0—Teachers express dissatisfaction with the quality and quantity of professional development opportunities.
1—Teachers just mention professional development opportunities.
2—Teachers mention professional development opportunities and discuss what they have learned from district workshops, research (e.g., CIERA website, journal articles) with other staff; there is some sense that teachers are trying to implement new ideas.
3—Professional development is ongoing; teachers have time to discuss, share, reflect on their practice, engage in professional development together across the building = collaborative learning community.
E. Leadership building:
0—Teachers express dissatisfaction with their school and the school's administration.
1—Teachers express dissatisfaction with their school or may be detached from the problems of their school without taking responsibility for implementing change; teachers express low to moderate satisfaction with the school administration.
2—Some teachers assume instructional leadership in the school; teachers express moderate to high satisfaction with school administration.
3—Includes those activities listed in E2 rating, as well as the following: Principal or administrative staff are strong leaders who also get teachers involved in leadership; time is provided for teachers to operate as a collaborative learning community; leadership helps the school use data to reflect on where they are and where they want to be (not just student assessment data, but current research on best practices); teachers express high satisfaction with school administration.

of the School Change Framework (see Table 2). We summed across the 10 components (for each component a point was awarded if the component was judged to be present) to generate a reform-effort score for each school. One member of the research team rated each school on each of the 10 components of implementing the reform. A second member of the research team also read through the artifacts and rated each school. The Pearson correlation coefficient across the two scorers' ratings was .92. The two raters met to discuss and resolve disagreements. Scores on the reform-effort scale ranged from a low of 1.0 to a high of 7.0, out of a possible

score of 10 ($M = 3.54$, $SD = 1.76$). This observed range of fidelity is consistent with other efforts of this sort (Bryk et al., 1998). More important, this variability, which we anticipated from the outset, allowed us to evaluate the relationship between reform effort and student growth.

Classroom practices

On three occasions (fall, winter, spring), each teacher selected to be part of the data collection sample was observed for an hour during reading instruction time to document curricular and pedagogical

TABLE 2
REFORM IMPLEMENTATION RUBRIC

One point was awarded for each of the reform components if the criteria in parentheses for a particular component were judged to be met.

1. Meeting for one hour three times per month in study groups (at least 80% of the time).
2. Meeting in cross-grade study groups (at least 80% of the time).
3. Reflecting on instruction and student work (demonstrated at least 80% of the time).
4. Considering research-based practices (demonstrated at least 80% of the time).
5. Being guided by action plans (yes or no).
6. Sticking with substantive topics for 3–4 months or more (yes or no).
7. Meeting once a month as a whole faculty to share and set goals (at least 80% of the time).
8. Working on a plan to involve parents as partners (yes or no).
9. Making effective use of an external facilitator (yes or no).
10. Making effective use of an internal leadership team (yes or no).

classroom practices in the teaching of reading. All observations were scheduled. The observers were retired teachers or graduate students in literacy education who had received training in the use of the CIERA Classroom Observation Scheme (Taylor & Pearson, 2000), and they were expected to demonstrate at least 80% agreement with a “standard” coding at each of the seven levels of the coding scheme prior to conducting classroom observations.

The observation system (influenced by the work of Greenwood, Carta, Kamps, & Delquadri, 1995; Scanlon & Gelzheiser, 1992; and Ysseldyke & Christenson, 1993–1996) combined qualitative note-taking with a quantitative coding process. The observer took field notes for a five-minute period, recording what was happening in the classroom, including, where possible and appropriate, what the teacher and children were saying. At the end of the note-taking period the observer recorded the proportion of children in the classroom who appeared to be on task (i.e., doing what they were supposed to be doing). The observer then coded the three or four most salient literacy events (Level 4 codes) that occurred during that five-minute episode. For every Level 4 event, the observer also coded who was providing the instruction (Level 1), the grouping pattern in use for that event (Level 2), the major literacy activity (Level 3), the materials being used (Level 5), the teacher interaction styles observed (Level 6), and the expected responses of the students (Level 7). An example of a five-minute observational segment is provided in Table 3 (see Table 4 for a list of the codes for all the levels). In Table 3, the codes “c/s/r” refer to levels 1–3, and codes “r/n/a/r,” “wr/n/c/or-tt,” and “/n/r/or” each refer to levels 4–7.

TABLE 3
SAMPLE OF OBSERVATIONAL NOTES

9:38—Small group continues. T is taking running record of child’s reading. Others reading familiar books. Next, T coaches boy on sounding out *discovered*. Covers up word parts as he says remaining parts. T: “Does that make sense?” T: “What is another way to say this part [‘cov’ with short ‘o’]”? T passes out new book: *My Creature*. T has students share what the word *creature* means. Ss: animals, monsters, dinosaurs, Dr. Frankenstein. 11/12 OT (On Task)
c/s/r r/n/a/r wr/n/c/or-tt v/n/r/or

Reliability of the observation codes

As the first author of this article visited research sites, she joined each observer in a 30-minute practice observation in order to establish interrater reliability data on the observation coding scheme. Across 12 abbreviated observations, agreements with the senior author were as follows: 95% at Level 2 (grouping), 95% at Level 3 (major literacy activity), 82% at Level 4 (specific literacy activity), 87% at Level 5 (material), 85% at Level 6 (teacher interaction style), and 82% at Level 7 (expected student response).

An expert observer, who had completed many classroom observations using this scheme and who had helped to refine it, read all of the observations to assess the degree to which observers were using the codes in a consistent manner. For example, although decision rules had been established in order to help observers distinguish between similar codes, one observer may have coded a teacher’s reference to the main idea of a story as a comprehension skill, while another observer might have coded a very similar exchange as a higher level question about the story. The expert observer did not code the observations “blind.” Instead, she recorded a different code only if she could not agree with the observer’s code after reading the narrative description of a particular five-minute segment. For a random sample of 10% of the observations, the agreements between the observers and expert observer at each of the levels of coding were measured as follows: 99% agreement at Level 2 (grouping), 100% at Level 3 (major literacy activity), 85% at Level 4 (specific literacy activity), 99% at Level 5 (material), 86% at Level 6 (teacher interaction styles), and 87% at Level 7 (expected student response). Because there was variability between the observers and the expert on this 10% sample, especially at Levels 4, 6, and 7, a decision was made to use the expert’s codes in all of the observations for those instances in which the observer and expert

Taylor et al., 2000; Wolf et al., 2000). Schools also received summary classroom observation data, school effectiveness data, and reform-effort data from schools in the project the previous year; the expectation was that they would use these findings as benchmarks for assessing their own strengths, weaknesses, and progress as the year unfolded.

At the beginning of a school's second year in the project, the school received a report that included information from the research on effective reform plus data on their classroom reading instruction (from the observations), ratings of school effectiveness (from the interviews), and reform effort (from the collection of artifacts described above). Schools were encouraged to interpret their classroom observation data in light of (a) the research on effective reading instruction, (b) the findings from the statistical analyses from the previous year(s) of the project, and (c) their observation data benchmarked against the entire sample of 13 schools. These analyses investigated the impact of various classroom practices on students' growth in reading achievement.

At the beginning of the second year in the project, schools were encouraged to interpret the school effectiveness data by comparing their mean ratings of shared leadership, collaboration, ongoing professional development, reflection on teaching, and parent partnership with the mean ratings for these five characteristics for all schools in the project. Schools were encouraged to reflect on their success in implementing the CIERA School Change Framework by considering whether each of the 10 elements in the reform rubric were rated as in place or not in place in their school.

Feedback to teachers

Throughout the year, teachers received copies of their observations, a description of the codes used in these observations, and a brief summary of research related to the major observation variables that were analyzed (e.g., incidence of higher level questioning, incidence of coaching). To help them interpret their own data they received a table summarizing observation data from teachers in the previous year. Teachers were encouraged to discuss their reports with their facilitators. External facilitators received training in how to interpret observations so that they, in turn, could help teachers understand the information contained in these observations. Facilitators, however, were directed not to interpret observations for teachers.

Design and data analyses

Hierarchical linear modeling (HLM; Bryk & Raudenbush, 1992) was used to analyze the impact of school-level and classroom-level characteristics on students' reading growth. The primary analyses in this study employed a three-level HLM model in which students were nested within classrooms and classrooms were nested within schools. In secondary analyses of the impact of reform effort on students' growth in reading over two years, HLM growth curve analyses were used. In these analyses, assessment data from the four time points were nested within students, and students were nested within schools. Descriptive analyses were also conducted to elaborate on the quantitative findings.

Three dependent measures—standardized comprehension, fluency, and writing scores—were analyzed independently using the HLM method. At the school level three predictor variables were used in the statistical analyses—school effectiveness score, reform-effort score, and year in study. At the classroom level, predictor variables from the classroom observations (i.e., those found to be important in previous research) were analyzed. These classroom practices included variables pertaining to grouping practices, literacy activities, text type, teacher response and student response. (see Table 5 for the variables and their descriptions).

HLM is a method of completing regression at multiple levels. It essentially estimates a regression within each classroom and school and combines these to see if they point to a common regression across classrooms and schools. When regressions (either the intercepts or slopes) vary across schools, we can examine the school-level or classroom-level characteristics that may explain such variation. This is a common method for evaluating school-level and classroom-level factors and their effects on student outcomes. A simple regression would be inappropriate in this situation, since it would assume observations to be independent, which is untenable in this situation because students in a given classroom are influenced by factors operating within that classroom.

HLM also partitions variance components across levels, providing an estimate of variance in student performance within and between classrooms and schools. An unconditional HLM is one without an explanatory variable that allows us to answer the following question: How much variance in student outcome can be attributed to systematic differences between classrooms and schools on specific factors? This analysis is equivalent to a random-effects analysis of variance.

TABLE 5
DESCRIPTION OF CLASSROOM OBSERVATION CATEGORIES USED IN DATA ANALYSIS

Percentage of time (five-minute segments) coded

Whole class or large group: All of the children in the class (except for one or two or individuals working with someone else) or a group of more than 10 children. If there are 10 or fewer in the room, code this as a small group.

Small group: Children are working in two or more groups. If there are more than 10 children in a group, call this whole group.

Narrative text: The number of segments in which a narrative textbook (tn) or narrative trade book (n) was coded out of the total number of segments coded.

Informational text: The number of segments that an informational textbook (ti) or informational trade book (i) was coded as being used out of the total number of segments coded.

Telling: Telling or giving children information, explaining how to do something.

Recitation: The teacher is engaging the students in answering questions, or responding, usually low-level q-a-q-a. The purpose primarily appears to be getting the children to answer the questions asked rather than engaging them in a formal discussion or fostering independence in terms of answering questions with more complete thinking.

Modeling: The teacher is showing/demonstrating how to do something or how to do a process as opposed to simply explaining it.

Coaching: The teacher is prompting/providing support that will transfer to other situations as students are attempting to answer a question or to perform a strategy or activity. The teacher's apparent purpose is to foster independence, to get a more complete thought or action rather than to simply get a student to answer a question.

Percentage of all reading segments coded

Phonemic awareness instruction: Students are identifying the sounds in words or blending sounds together (an oral activity). The purpose is to develop phonemic awareness, not letter-sound knowledge (e.g., Sound Box technique would be coded as "pa" since the focus is on learning the sounds in words).

Phonics instruction: Students are focusing on symbol/sound correspondences (p1) or letter-by-letter decoding (p2) or decoding by onset and rime or analogy (p3), but this is not tied to decoding of words while reading. If students are decoding multisyllabic words, code as p4. The total number of phonics activities out of total number of times reading was coded at level 3 was calculated.

Word recognition strategies: Students are focusing on use of one or more strategies to figure out words while reading, typically prompted by the teacher.

Lower level text comprehension (talk or writing about text): Students are talking (m1) or writing (m2) about the meaning of text that is at a lower level of thinking. The writing may be a journal entry about the text or a fill-in-the blank worksheet that is on the text meaning (rather than on a comprehension skill or vocabulary words). The total number of "low-level text comprehension" activities at level 4 out of the total number of times reading was coded at level 3 was calculated.

Higher level text comprehension (talk or writing about text): Students are talking (m3) or writing (m4) about the meaning of text that is engaging them in higher level thinking. This is talk or writing about text that is challenging to the children and is at either a high level of text interpretation or goes beyond the text: generalization, application, evaluation, aesthetic response. Needless to say, a child must go beyond a yes or no answer (e.g., in the case of an opinion or aesthetic response). The total number of "high-level text comprehension" activities at level 4 out of the total number of times reading (as the major focus) at level 3 was coded.

Comprehension skill instruction: Students are engaged in a comprehension activity (other than a comprehension strategy) that is at a lower level of thinking (e.g., traditional skill work such as identifying main idea, cause-effect, fact-opinion).

Comprehension strategy instruction: Students are using a comprehension strategy that will transfer to other reading and in which this notion of transfer is mentioned (e.g., reciprocal teaching, predicting; if predicting were done, but transfer was not mentioned, this would be coded as "c").

Vocabulary instruction: Students are discussing/working on a word meaning(s).

Active reading practice: Students are reading (not reading turn-taking) at level 7.

Percentage of all codes for student responding

Active responding: Children are engaged in one or more of the following level 7 responses: reading, writing, oral responding, manipulating. The total number of "active responding" codes coded out of the total number of level 7 responding codes coded was calculated.

Passive responding: Children are engaged in one or more of the following level 7 responses: reading turn-taking, oral responding turn-taking, listening. The total number of "passive responding" codes coded out of the total number of level 7 responding codes coded was calculated.

Time on task: At the end of the five-minute note-taking segment, the observer counted the number of children in the room who appeared to be engaged in the assigned task out of all the children in the room. If a child was quiet but staring out the window or rolling a pencil on his or her desk, this was not counted as on task.

HLM is a tool for testing theory-based models and for explaining variation in outcome measures. In this article, we focus on the amount of variation explained and effect size for relevant variables rather than a significance standard for exploring issues related to classroom and school practices and gains in

reading ability (Lee & Smith, 1997). We interpreted effect sizes as follows: 0.1–0.3, small; 0.3–0.5, moderate; and 0.5 or more, large (Rosenthal & Rosnow, 1984). We extended the conventional *p*-value decision rule beyond 0.05, consistent with the practice of current HLM researchers by retaining effects up

TABLE 6
MEANS AND STANDARD DEVIATIONS FOR STUDENT SCORES GRADES 2–5

Assessment tool/grade	N	Fall		Spring	
		M	SD	M	SD
Fluency					
Grade 2	174	64.07	35.38	81.53	33.07
Grade 3	200	87.11	33.50	92.12	32.94
Grade 4	183	100.10	35.03	121.99	41.82
Grade 5	176	125.18	37.83	135.16	38.96
Gates–MacGinitie comprehension (NCE)					
Grade 2	169	44.43	17.78	44.11	18.90
Grade 3	199	39.60	17.25	41.02	17.10
Grade 4	180	35.49	17.37	36.58	18.06
Grade 5	175	38.16	16.20	38.54	16.87
Writing					
Grade 2	152	1.49	.70	1.86	.79
Grade 3	169	1.43	1.55	1.53	.60
Grade 4	159	1.37	.51	1.64	.67
Grade 5	127	1.64	.65	1.76	.77

to $p < 0.10$ when it is theory based and there is prior evidence in the literature regarding the effect (Lee & Bryk, 1989; Lee & Smith; Mayer, 1998), and when it explains substantial variation (Raudenbush & Bryk, 2002; Simpson, Raudenbush, & Earls, 1997; Wenglinsky, 1998). For a more complete description of estimation in HLM, see Bryk and Raudenbush (1992, pp. 32–56). *HLM* (Raudenbush, Bryk, Cheong, & Congdon, 2000) is recognized as a stan-

dard program for estimating multilevel models (Bryk & Raudenbush; Kreft & DeLeeuw, 1998).

Results

Results were analyzed across grades 2–5 since students at these grades had the same measures in the fall and spring. Student scores are in Table 6. Classroom practices by grade are in Table 7.

TABLE 7
INCIDENCE OF CLASSROOM FACTORS BY GRADE

	N =	Grade 2		Grade 3		Grade 4		Grade 5		Mean across grades	
		M	SD	M	SD	M	SD	M	SD	M	SD
		23		24		23		22		92	
<i>Percentage of segments coded out of all segments coded</i>											
Whole group		42	38	33	32	40	32	49	25	40	32
Small group		50	37	64	29	58	32	55	32	56	32
Informational text		9	19	23	22	17	22	25	22	19	22
Narrative text		64	23	51	24	66	22	44	30	57	26
Telling		60	19	64	17	64	19	73	18	65	18
Recitation		74	17	72	19	74	18	70	21	73	18
Coaching		29	21	18	13	18	18	19	20	21	19
Modeling		8	13	6	9	8	9	7	8	7	10
<i>Percentage of segments coded out of all reading segments</i>											
Phonemic awareness		5	7	1	5	—	—	—	—	—	—
Phonics instruction		11	12	4	8	6	2	4	9	5	9
Word recognition strategies		15	14	8	8	7	10	6	9	9	11
Vocabulary		29	18	27	19	29	16	29	20	28	18
Comprehension skills		12	10	18	17	13	12	17	17	15	14
Comprehension strategies		4	11	5	7	6	13	6	10	5	11
Meaning of text—lower level		38	19	58	28	53	20	46	29	49	25
Meaning of text—higher level		10	11	20	18	22	15	22	22	18	18
<i>Percentage of responses coded out of total number of Level 7 responses</i>											
Active Responding		42	10	37	11	31	8	33	11	36	11
Passive responding		58	10	63	11	69	8	67	11	64	11

Standardized comprehension scores

From the three-level HLM analysis (HLM3; Raudenbush et al., 2000) on Gates–MacGinitie comprehension normal curve equivalent (NCE) scores, after accounting for fall scores, we found that 24% of the variance was between teachers and 10% of the variance was between schools. (See Table 8 and Appendix.)

Reform-effort rating was positively related to students' spring standardized reading comprehension scores, accounting for 17% of the between school variance ($ES = .29$). For every 1-point increase in reform-effort score, a school's mean NCE score increased by 1.34.

At the classroom level, we found that grade ($ES = .36$) and the coding of comprehension skill instruction ($ES = .27$), both negatively related, accounted for 29% of the between teacher variance. For every increase in grade level (e.g., from grade 2 to 3 or 3 to 4), students' mean NCE score decreased by 2.57. For every 10% increase in the coding of comprehension skill practice, a students' mean NCE score decreased by 1.38.

Fluency scores

When considering students' fluency scores, after accounting for fall scores and grade, 19% of the variance was between teachers and 22% between schools. Reform effort accounted for 35% of the be-

tween school variance ($ES = .38$). For every 1-point increase in reform-effort score, students' mean oral reading fluency (wcpm) score increased by 4.87. High-level questioning (positively related, $ES = .27$) and comprehension skill instruction (negatively related, $ES = .31$) accounted for 15% of the between teacher variance. For every 10% increase in the coding of higher level questioning within a classroom, students' mean fluency score increased by 1.75. For every 10% increase in the coding of comprehension skill practice within a classroom, students' mean fluency score decreased by 2.62 (see Table 9).

Writing

When considering students' writing scores, 32% of the variance was between teachers and 4% between schools. No school-level variables were found to contribute to the between school variance. Coaching (positively related) accounted for 11% of the between teacher variance ($ES = .38$). For every 10% increase in the coding of coaching within a classroom, students' mean writing score (based on a 4-point rubric) increased by 0.08 (see Table 10).

Growth curve analysis to investigate change in students' performance

Eight schools were in the project across the two years, and approximately one third of the students

TABLE 8
GRADES 2–5 READING COMPREHENSION

Initial random effects	Variance component	% variance between		
Classroom means	49.72	24*		
Student residual	141.11			
School means	20.68	10		
Classroom fall score slope	.03			
Total	211.54			
Final random effects		% variance accounted for by model		
Classroom means	35.15	29*		
Student residual	141.63			
School means	17.16	17*		
Classroom fall score slope	.026			
Final fixed effects	Coefficient	<i>t</i> ratio	<i>df</i>	<i>p</i> value
Intercept (Grand mean)	39.57	28.54	11	.000
Reform effort (school)	1.34	1.80	11	.098
Grade (classroom)	–2.57	–3.63	88	.001
Comprehension skills (classroom)	–13.78	–2.29	88	.022
Fall score (student)	.67	19.09	90	.000

*Variance between classrooms = $49.72/211.54$. Classroom-level variance accounted for by the model = $(49.72 - 35.15)/49.72$. School-level variance accounted for by the model = $(20.68 - 17.16)/20.68$.

TABLE 9
GRADES 2–5 READING FLUENCY

Initial random effects	Variance component	% variance between		
Classroom means	144.92	19		
Student residual	412.10			
School means	163.69	22		
School grade slope	24.43			
Classroom fall score slope	0*			
Total	745.14			
Final random effects		% variance accounted for by model		
Classroom means	122.59	15		
Student residual	412.25			
School means	105.71	35		
School grade slope	27.28			
Final fixed effects	Coefficient	<i>t</i> ratio	<i>df</i>	<i>p</i> value
Intercept (Grand mean)	106.41	33.49	11	.000
Reform effort (school)	4.87	3.90	11	.003
Grade (classroom)	17.97	9.11	12	.000
High-level questioning (classroom)	17.49	1.81	88	.070
Comprehension skills (classroom)	–26.26	–2.35	88	.019
Fall score (student)	.82	34.14	727	.000

*This variable was nonrandomly varying and thus fixed in the model.

across eight schools were in the study for two years. By employing the four time points across two years (fall year 1, spring year 2, fall year 1, and spring year 2), a three-level HLM was fit to the data; time points were nested within students, and students were nested within schools. This allowed us to estimate an intercept (performance level in fall of year 2) and a slope (growth rate across the four time points). The growth curve analyses yielded significant results for comprehension and fluency, but not for writing.

For Gates–MacGinitie comprehension results with students in grades 2–5, an unconditional model estimated the intercept at 41.7 NCEs (fall year 2) with an average slope of 0.23 NCEs per time point (see Table 11). (Because of some missing data, which HLM can accommodate, we chose fall of year 2 as the centering point for the intercept.) Both the intercept and slope of the growth curves varied significantly between schools. For reading comprehension we found that 6% of the variance in fall year 2 status was

TABLE 10
GRADES 2–5 WRITING

Initial random effects	Variance component	% variance between		
Classroom means	.164	32		
Student residual	.304			
School means	.023	4		
School fall score slope	.025			
Total	.516			
Final random effects		% variance accounted for by model		
Classroom means	.146	11		
Student residual	.304			
School means	.025	0		
School fall score slope	.025			
Final fixed effects	Coefficient	<i>t</i> ratio	<i>df</i>	<i>p</i> value
Intercept (Grand mean)	1.69	24.64	10	.000
Coaching (classroom)	.80	2.39	76	.017
Fall score (student)	.28	4.04	10	.003

between schools and 15% of the variance in growth across two years was between schools. Grade had a significant relationship with student intercepts, where higher grades performed at a slightly lower level in terms of NCE scores (-3.03 NCEs, $p = 0.001$). This finding is similar to recent findings from California showing that the longer our most vulnerable students stay in school, the worse they perform relative to national norms (Calfee, 2003). Reform-effort scores explained a significant amount of variation in growth curve slopes (65% of the variation between schools was explained, $ES = .49$; see Table 11). On average, 1 point on the reform-effort scale increased the growth slope by 0.63 NCEs per time point; over two years, 1 additional point in reform effort is associated with 2.5 NCEs additional growth.

A similar result was found with the reading fluency measure (see Table 12). The average intercept (average wcpm fall year 2) was 97.2 with an additional 20.8 wcpm for each additional year in grade (from 2 to 5). The average growth slope was 12.5 per time point with a slight decrease in growth rate for each additional grade of 1.6 wcpm (fluency in earlier grades grows faster). Both the intercept and growth slope varied significantly between schools. Reform-effort scores were a reasonable explanatory variable, indicating an increase in wcpm of 0.78 per time point; in two years, 1 additional point in reform effort is associated with 3.1 wcpm growth in addition to the mean growth slope of 12.5 wcpm ($p = 0.07$, $ES = .41$; see Table 12).

The estimated variance appears to have increased in the final model, suggesting that the addition of the reform-effort score created greater spread in school growth rates; however, reform effort was modestly statistically significant in the model ($p = 0.074$). In addition, the model fit index suggested that the inclusion of reform effort improved the fit to the data (chi-square = 3.9, $df = 1$, $p < 0.05$).

Summarizing across HLM findings

The HLM3 analyses highlight both school- and classroom-level factors in explaining students' reading achievement. Both levels contributed a moderate amount to students' growth in reading. For comprehension, 24% of the variance in students' growth was between teachers and 10% was between schools. For fluency, 19% of the variance in students' growth was between teachers and 22% was between schools.

Looking across a single school year, the impact of reform effort was relatively small when considering effect size. This is not surprising since it is well established that reform is a gradual process. When looking at the impact of reform effort across two years, however, we found a much larger impact; 65% of the variance between schools was accounted for by reform effort. Across two years, a school that had 1 additional point on the reform-effort scale had 2.5 NCEs of growth in reading comprehension above

TABLE 11
GRADE 2-5 GROWTH CURVE ANALYSIS—READING COMPREHENSION

Initial random effects	Variance component	% variance between		
Student status fall year 2	190.04			
Student growth slope	9.49			
Student residual	88.37			
School status fall year 2	17.93	6*		
School growth slope	1.66	15*		
Total	296.34*			
Final random effects		% variance accounted for by model		
Student status fall year 2	179.83			
Student growth slope	9.36			
Student residual	88.25			
School status fall year 2	18.71			
School growth slope	.576	65*		
Final fixed effects	Coefficient	<i>t</i> ratio	<i>df</i>	<i>p</i> value
School status	41.74	22.87	7	.000
Grade	-3.03	-3.64	239	.001
School growth	.23	.47	6	.652
Reform effort	.63	3.00	6	.026

*Total = 190.04 + 88.37 + 17.93. Variance in status between schools = 17.93/296.34. Variance in growth between schools = 1.66/(9.49 + 1.66). Percentage of variance in growth slope accounted for by model = (1.66 - .576)/1.66.

TABLE 12
GRADES 2–5 GROWTH CURVE ANALYSIS—READING FLUENCY

Initial random effects	Variance component	% variance between		
Student status fall year 2	717.75			
Student growth slope	16.09			
Student residual	290.25			
School status fall year 2	162.90	14		
School growth slope	3.56	18		
Total	1170.90			
Final random effects		% variance accounted for by model		
Students status fall year 2	713.55			
Student growth slope	15.17			
Student residual	290.49			
School status year 2	168.50			
School growth slope	4.35			
Final fixed effects	Coefficient	<i>t</i> ratio	<i>df</i>	<i>p</i> value
School status	97.20	19.45	7	.000
Grade	20.83	12.57	239	.000
School growth	12.52	12.38	6	.000
Grade	−1.57	−2.52	239	.012
Reform effort	0.78	2.16	6	.074

and beyond a school with 1 point less on the reform-effort scale.

Our reform-effort scale corroborates and extends earlier work on school improvement and effective schools. As in earlier work, we found a consistent cluster of influential practices—teachers learning and changing together over an extended period of time, reflection and dialogue on practice and implementation of research-based teaching strategies (Bryk et al., 1998; Fullan, 1999; Langer, 2000; Louis & Kruse, 1995), and collaborative leadership (Wolf et al., 2000).

Looking at the classroom level across grades 2–5, a number of findings converge with our earlier research as well as the research of others. We found that higher level questioning contributed to the between teacher variance in students' fluency scores in grades 2–5, whereas rote comprehension skill practice (which was coded separately from comprehension strategy instruction) was negatively related to both reading comprehension and fluency growth in grades 2–5. Similar findings on the importance of higher level questioning were reported in our earlier related studies (Taylor et al., 2000; Taylor et al., 2003; Taylor, Peterson, et al., 2002) as well as in other research (Knapp, 1995). In a related paper that focused more exclusively on classroom-level analyses (Taylor et al., 2003), we have argued that the combination of higher level activities and teacher practices that support active and sustained student learning

can be viewed as supporting a stance that we have labeled teaching for cognitive engagement. When viewed in the context of other efforts, such as those of Langer (1999, 2000), Allington and Johnston (2002), and Knapp, the evidence for emphasizing meaning and higher order processes with low-income populations seems impressive.

The National Reading Panel found that comprehension strategy instruction, as opposed to comprehension skill practice, was important for students' reading growth. However, we saw such low levels of comprehension strategy instruction (see Table 7) that it is not surprising it did not emerge as a significant factor in the HLM analyses conducted. We also found in earlier work that a relatively high level of comprehension skill practice was negatively related to reading comprehension growth in grades 2–5 (Taylor et al., 2003). A logical explanation is that high amounts of mechanistic practice on comprehension skills are taking time away from other important comprehension activities such as higher level talk about text and use of comprehension strategies during reading.

Coaching was positively related to writing growth in grades 2–5. This finding is related to earlier work in which coaching was found to benefit students' reading growth (Taylor et al., 2000; Taylor et al., 2003). The importance of coaching has also been highlighted by Allington and Johnston (2002) and Pressley et al. (2001).

Differences in reform effort across schools

Reform effort had a small but significant impact on growth in school performance when considering scores from fall to spring in the same school year. However, when looking at schools that had been in the project for two years, the reform effort had a relatively large impact on students' reading growth, especially in reading comprehension. To more fully understand the impact of reform effort, we examined differences across all 13 high-reform-effort (HRE) and low-reform-effort (LRE) schools in implementation of the reform and in perceptions of school effectiveness. We also examined the eight HRE and LRE schools that had been in the project for at least two years in order to study changes in teachers' classroom practices.

First we categorized schools as high-, medium-, or low-reform-effort schools. The mean reform-effort score was 3.7 with a standard deviation of 1.9. The five schools that achieved a reform-effort rating score of 5, 6, or 7 (more than one half a standard deviation above the mean) were designated HRE schools. Three schools earned a reform-effort rating of 4 (less than half a standard deviation above or below the mean) and were designated moderate-reform-effort schools. The five schools with a reform-effort rating of 1 or 2 (more than one half a standard deviation below the mean) were designated as LRE schools. The high-, medium-, and low-reform-effort schools exhibited only modest variation in terms of percentage of students on subsidized lunch (the means were 81%, 73%, and 87%, respectively), of ELL (the mean percentages were 16, 16, and 25, respectively), or students of color (the mean percentages were 62, 75, and 78, respectively).

Of the eight schools in the project for two years, four were high-reform, one was moderate-

reform, and three were low-reform schools. When examining the students who had been assessed for two years in the high- and low-reform-effort schools, we found that the mean spring Gates–MacGinitie comprehension NCE score for students in the high-reform schools increased from year 1 to year 2 (except in grade 2) whereas the mean score for students in the low-reform-effort school actually decreased (see Table 13).

Differences in reform efforts across high- and low-reform-effort schools

On the basis of the reform-effort ratings (see Table 14), we conducted *t* tests comparing the number of HRE and LRE schools determined to be engaging in various reform practices, setting the alpha level at .01 because multiple *t* tests were performed (we followed this procedure throughout this section). Analyses revealed that more of the HRE schools than LRE schools engaged in these school-level practices: sticking with a substantive study-group topic for at least three to four months, $t(8) = 4.00, p = .004$; meeting once a month to share study-group activities, $t(8) = 4.00, p = .004$; and having an effective internal leadership team, $t(8) = 4.00, p = .004$.

Study-group topics identified as *substantive* included research-based reading practices shared with teachers at the beginning of the school year. For example, teachers were encouraged to increase higher level questioning, comprehension strategies, or the application of phonics to the reading of connected text (also see under the heading "Data for the schools"). In addition, the HRE schools were meeting regularly as a large group to share study-group successes and to deal with schoolwide issues pertaining to literacy.

TABLE 13
STANDARDIZED SPRING COMPREHENSION SCORES IN HIGH- AND LOW-REFORM SCHOOLS FOR STUDENTS IN FIRST AND SECOND YEAR IN STUDY

Grade in year 2 for students in HRE or LRE schools	<i>N</i>	Year 1 spring Gates–MacGinitie NCE		Year 2 spring Gates–MacGinitie NCE	
		<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
2 HRE	20	52.10	16.47	48.00	16.78
2 LRE	16	47.44	16.07	45.41	9.65
3 HRE	33	44.77	15.43	46.14	17.83
3 LRE	11	32.14	29.44	27.82	8.15
4 HRE	28	36.73	14.00	38.89	14.41
4 LRE	17	36.11	10.78	32.18	16.11
5 HRE	28	40.02	15.50	42.20	13.77
5 LRE	17	35.94	18.29	35.09	13.89

TABLE 14
REFORM-EFFORT RATINGS

Reform-effort variable	Percentage of HRE schools ($n = 5$) demonstrating this reform variable	Percentage of LRE Schools ($n = 5$) demonstrating this reform variable
Meeting for one hour three times per month in study groups	80	20
Meeting in cross-grade study groups	40	20
Reflecting on instruction and student work	60	0
Considering research-based practices	40	0
Being guided by action plans	20	0
Sticking with substantive topics for three to four months or more	80*	0
Meeting once a month as a whole faculty to share, etc.	80*	0
Working on a plan to involve parents as partners	40	0
Effective use of external facilitator	40	60
Effective use of internal leadership team	100*	20

* $p = .004$

Leadership teams rated as effective were typically led by a teacher leader who was (a) very knowledgeable about reading and (b) respected by the other teachers. Other members of effective leadership teams encouraged teachers to continue to meet in study groups, helped to run study-group meetings, and met regularly to discuss the progress of study groups and solve problems. Also, the high-reform-effort schools typically had a supportive principal who was enthusiastic about the reform effort.

Although all HRE schools possessed effective internal leadership teams, only two schools had an external facilitator who worked at the school regularly (in spite of the recommendation that such a person be in place). One school did not have the funds for an external facilitator, and two schools that shared an external facilitator saw relatively little of this person because she had been assigned by the district to work with two other schools as well.

School effectiveness rating across high- and low-reform-effort schools

The school effectiveness rating did not enter into any of the HLM3 models as a school-level factor contributing to students' literacy growth. However, an independent t test revealed that teachers in HRE schools had more positive comments about their professional development than teachers in LRE schools, $t(9) = 3.67$, $p = .01$ (see Table 15). This is not surprising because teachers in the HRE schools were sticking with substantive topics over time in study groups, whereas in the LRE schools teachers were not. It is likely that teachers in the HRE schools felt their study-group work was sustained and valuable, whereas teachers in the LRE schools may have sensed that their study-group work was unfocused or on unsubstantial topics. That being said, the mean ratings for professional development and reflection on practice in study groups in the HRE schools were 1.9 (SD

TABLE 15
SUMMARY DATA FROM THE TEACHER INTERVIEWS AND DESCRIPTIONS
OF CATEGORIES ANALYZED

Teacher perceptions	Mean rating for HRE schools (based on 4-point rubric, where 0 = low and 3 = high)		Mean rating for LRE schools	
	M	SD	M	SD
Links to parents	1.40	.45	1.55	.32
Collaboration	1.83	.24	1.46	.73
Professional development	1.90*	.18	1.41	.27
Reflection on teaching	1.87	.27	1.37	.43
Collaborative leadership	1.72	.32	1.44	.43
Total	8.72	1.32	7.22	2.03

* $p = .01$

= 0.2) and 1.9 ($SD = 0.3$), respectively, on a scale ranging from 0 to 3. This suggests that even in the schools doing the best job of implementing study groups, there was still more that schools could do to become collaborative learning communities in which teachers were reflecting on practice and working together to improve instruction.

Changes in teaching practices across high-reform-effort and low-reform-effort schools

Because we had decided a priori that three observations per year did not provide enough data to examine within-year changes in instructional practice, we examined cross-year changes in the HRE schools ($n = 4$) and LRE schools ($n = 3$) that had been in the reform effort for two years. Unfortunately, only some of the same teachers within these schools were observed in both year 1 and year 2. Thus, considering all teachers in year 1 and year 2 in two-year schools, we were unable to statistically compare changes in teaching practices within HRE and LRE schools across the two years.

However, we were able to look at the teaching practices of a subset of teachers in HRE and LRE schools that had actually been observed in both year 1 and year 2 (see Table 16). Using paired t tests for each group, we found that the HRE teachers increased their use of coaching from year 1 to year 2, $t(17) = 3.46$, $p = .003$. No significant differences from year 1 to year 2 were found for the LRE teachers.

We also considered differences between HRE and LRE teachers in year 1 and HRE and LRE teachers in year 2. Using independent t tests, we found that teachers in HRE schools were observed asking significantly more high-level questions in year 2, $t(50) = 2.62$, $p = .01$, than teachers in LRE

schools. Teachers in HRE schools were observed doing significantly more modeling in year 2 than teachers in LRE schools, $t(50) = 3.54$, $p = .001$ (see Table 17).

Collectively, these findings suggest that compared to the teachers in LRE schools, the teachers in the HRE schools were making more of an effort to examine the data on their teaching practices and to use effective teaching practices or to change their reading instruction in the directions suggested by the research. However, in neither high- nor low-reform-effort schools did the incidence of comprehension strategies instruction increase. This latter finding is somewhat puzzling, given the research on the importance of comprehension strategies instruction and the apparent willingness of teachers in many schools to consider teaching them. One explanation is that comprehension strategies instruction is difficult to implement (Pressley, 2002).

A description of one high-reform-effort school

To provide a more vivid picture of what reform looked like in these schools, we describe the process of one high-reform-effort school, Howard Elementary (a pseudonym), over the two years of the project. Howard Elementary School was in a large U.S. urban area in which 81% of the students qualified for subsidized lunch and 78% of the students were English-language learners. We offer this description to illustrate how the study-group process, along with the use of student data, may have contributed to the changes in classroom teaching practices that were observed. This in turn may have led to the increased reading growth at Howard from the first year of the project to the second. Our language

TABLE 16
MEAN INCIDENCE OF CLASSROOM FACTORS* BY HRE AND LRE SCHOOLS AND YEAR
IN STUDY FOR GRADE 2-5 TEACHERS OBSERVED FOR TWO YEARS

	HRE year 1		HRE year 2		LRE year 1		LRE year 2	
	$N = 18$		$N = 18$		$N = 10$		$N = 10$	
	M	SD	M	SD	M	SD	M	SD
High-level questioning	.19	.15	.28	.26	.10	.12	.15	.23
Comprehension strategies	.08	.09	.05	.12	.04	.04	.02	.05
Vocabulary	.26	.23	.32	.15	.20	.17	.26	.17
Word recognition strategies	.08	.07	.09	.11	.16	.14	.10	.14
Coaching	.11	.12	.27**	.18	.11	.13	.18	.21
Modeling	.03	.03	.04	.06	.04	.07	.02	.04
Active responding	.29	.11	.36	.12	.34	.24	.29	.11

*Research shared with teachers recommended increasing incidence of this practice. **HRE year 2 > LRE year 2, $p = .003$.

TABLE 17
MEAN INCIDENCE OF CLASSROOM FACTORS* FOR ALL TEACHERS OBSERVED BY HRE
AND LRE SCHOOLS AND YEAR IN STUDY FOR GRADES 2-5

	HRE year 1		HRE year 2		LRE year 1		LRE year 2	
	N = 31		31		23		21	
	M	SD	M	SD	M	SD	M	SD
High-level questioning	.21	.16	.25**	.22	.11	.14	.11	.15
Comprehension strategies	.10	.15	.06	.12	.08	.10	.04	.07
Vocabulary	.27	.20	.28	.17	.30	.29	.24	.17
Word recognition strategies	.10	.13	.07	.09	.08	.10	.10	.14
Coaching	.12	.06	.24	.03	.18	.15	.16	.11
Modeling	.04	.05	.07***	.08	.08	.07	.01	.03
Active responding	.28	.12	.36	.11	.38	.20	.36	.13

*Research that was shared with teachers recommended increasing incidence of this practice. **HRE year 2 > LRE year 2, $p = .01$. ***HRE year 2 > LRE year 2, $p = .001$.

is intentionally cautious and circumspect because we are mindful of the difficulty in attributing causal connections between changes in practice and changes in student performance. Plausibility, not causality, is the goal of this description.

During the first year in the project, the teachers at Howard selected study groups, most of which, but not all, were directly influenced by the research report on effective reading instruction the school had received at the beginning of the year. Study groups focused on the following topics: guided reading, reading comprehension, reading assessment, reading interventions within the classroom, higher level questioning, and refining coaching and modeling abilities. All of the topics, except for the last two, were very broad. Also, meeting notes and facilitator log entries suggested that learning how to be productive in study groups took up a fair amount of teachers' energy in the first half of the first year.

During the second year of the project, as compared to the first year, teachers at Howard were more focused on specific instructional strategies to improve reading comprehension because their school report and district-level data indicated that this was their biggest challenge area. Teachers spent the first half of the year in cross-grade study groups learning how to teach children to use thinking maps to summarize what they read. During the second half of the year teachers met in cross-grade study groups that focused on additional strategies to improve students' comprehension. For example, one group refined its use of the Directed Reading-Thinking Activity (DRTA) routine (Blachowicz & Ogle, 2001). Another group learned how to teach students to use Students Achieving Independent Learning (SAIL; Brown, Pressley, Van Meter, & Schuder, 1996). One

group worked on developing challenging independent seat-work activities to foster reading comprehension, while another focused on vocabulary instruction to improve reading comprehension.

Studying the classroom teaching practices of the five teachers at Howard who were observed in both year 1 and year 2, we found that they made changes in their teaching practices in the directions suggested by the research (see Table 18). When considering these five teachers, we found that 3 of 5 were observed doing more high-level questioning, 2 of 5 increased their focus on comprehension strategies instruction, 2 of 5 increased their use of coaching, and 4 of 5 engaged students in more active responding in year 2 as compared to year 1.

Looking at two teachers

To illustrate teachers' changes in teaching practices at Howard across the two years of the project, we focus on two of the five teachers observed in both years. Mrs. Lopez (Teacher B; all names are pseudonyms) was a second-grade teacher in the higher level thinking study group during the first year in the project and in a thinking map study group and the SAIL study group during the second year. We contrast two observations, one from the fall of year 1 with a second from the spring of year 2. In fall of the first year as she was reading with a group of four students, she would stop at predetermined places in the story that she had marked for the children with sticky notes. Typically, her questioning was at a fairly low level. "Why is Joe so surprised? How do you know that? What else was he surprised about? What happened at night?" Then the teacher had the students continue reading.

TABLE 18
PERCENTAGE OF SEGMENTS IN WHICH TEACHING PRACTICE CHANGES WERE MADE BY
TEACHERS AT HOWARD WHO WERE OBSERVED FOR TWO YEARS

	Teacher									
	A—grade 2		B—grade 2		C—grade 3		D—grade 3		E—grade 5	
	Year		Year		Year		Year		Year	
	1	2	1	2	1	2	1	2	1	2
Higher level questioning	15	83	29	8	9	9	0	63	28	47
Comprehension strategies	11	0	19	46	9	0	27	13	0	5
Vocabulary	22	67	24	8	100	45	59	38	17	23
Word recognition strategies	7	17	24	0	9	9	18	13	0	0
Coaching	21	38	33	54	0	6	13	38	0	13
Modeling	2	0	6	0	4	6	0	13	11	13
Active responding	39	52	49	51	15	19	40	41	27	40

In spring of the second year Mrs. Lopez's questioning routine with a small group looked very different, and she emphasized the use of reading strategies. Also, the students were doing more of the work for themselves in the second year, largely because Mrs. Lopez emphasized recitation less and coaching more. For example, as she was working with a small group, the students started their reading of a new story about spiders by doing a picture walk on their own. Then, after they chorally read the first page, they each completed a story map independently with support from their teacher. At the end of the second page a student, without being prompted by the teacher, pointed out that there wasn't a problem so far to put on the map; just the characters and setting had been described. The students continued reading on their own, and then, as a group, they identified the problem of the story. At the end of the lesson the teacher reminded students that a story map helped them remember the important parts of a story and that they could use the strategy when they were reading on their own. As the group went back to their seats, a child proudly came up to show the teacher a strategy he had used while reading on his own. He had written this strategy on a sticky note so he wouldn't forget to share it with his teacher.

Ms. Gray (Teacher D) was a third-grade teacher in the reading comprehension study group during the first year and the thinking map and SAIL study groups in the second. During the beginning of the first year, Ms. Gray's lessons were fairly teacher directed and focused on low-level thinking. For example, in the fall of the first year, a small group was reading "Goldilocks and the Three Bears." Ms. Gray asked questions about the story, such as "What happened to the rocking chair? How did Mama feel when she saw the rocking chair?" (A student answered, "Bad") and "How would you describe the

bear?" The teacher did not ask students to elaborate on their brief answers, and thus the questions did not engage the students in more elaborate or higher level thinking about the text. Students then continued reading. In the winter of year 1 Ms. Gray and a group read a nonfiction story about penguins. Ms. Gray listed things the students told her they had learned about penguins. "What do penguins have? What do they eat?" At the end of the lesson Ms. Gray asked students to review with her what they had learned that day. "We worked on finding the meaning of a word. We read the table of contents. We learned about reading nonfiction books."

During the spring of the second year, Ms. Gray's small-group lessons looked very different. She included many more high-level questions than she had the year before. As she worked with one group, the students interpreted characters in the story they were reading. Ms. Gray asked, "What does Mrs. Gorf think of kids?" A student replied, "She thinks they are a bother." After reviewing the plot, Ms. Gray then asked, "What do you think is the theme of the story?" After students took turns sharing about the importance of being nice to people, Ms. Gray asked, "How does the author's message affect your life?" Students talked about things that someone might say or do that could hurt other people. Ms. Gray had students summarize what they had learned that day. "We learned about theme. We learned that it's important to be nice to other people." Students then went to their seats to work on a story map for the story they had just read and discussed.

Teachers' perceptions of important school factors at Howard

An analysis of grade K–5 teachers' comments on interviews (Taylor & Pearson, in press) revealed

striking consistencies across teachers at Howard. These comments parallel the findings of effective schools and effective school improvement reported in the research literature. Specifically, three types of responses illuminate the general findings on the reform effort: the importance of focus, the value of the study-group process, and the usefulness of the observation data.

Teachers at Howard saw improved reading comprehension for students as a personal and schoolwide goal, and clearly reading comprehension instruction was a major focus in grade 1–5 classrooms. Ten of 10 teachers interviewed in grades 1–5 talked about some aspect of reading comprehension when asked to describe three critical components of their reading program.

Across grades K–5, 10 of 12 teachers interviewed also mentioned that improved reading comprehension was a schoolwide goal. Most mentioned that there was consistency in their practices. They were all using thinking maps and writing in response to reading. They were using guided reading consistently, with a focus on higher level questions and use of comprehension strategies.

When asked about helpful opportunities for learning about literacy instruction, 12 of 12 teachers in grades K–5 made positive comments about the CIERA study-group process. The model was helping to provide consistency in instruction. Teachers were clear about the study-group process. Teachers uniformly commented on the value of reflection. Teachers also valued learning from one another.

Teachers talked about the value of change. Twelve of 12 teachers in K–5 reported that their teaching had changed in positive ways. Many teachers mentioned that they were now more confident about their teaching of reading. As one teacher commented, “I now see more clearly how to help kids, how to meet their individual needs.”

Six of 12 teachers mentioned the value of the collegiality the study-group process provided. One teacher summed it up well, “At first I thought study groups were a waste of time. But it’s been very positive. It has helped us use similar approaches. We have more unity. I hope we can continue with study groups. As teachers, we need time to reflect on the effectiveness of our lessons.”

Seven of 12 teachers interviewed mentioned the usefulness of the observation data to help them reflect on their teaching. Teachers made comments like the following:

The observation feedback has been useful in that it provides another pair of eyes.

The observation feedback helped me be more aware of the instructional strategies I used and to what extent my students were actively involved.

At first I was hesitant, but I’m glad for the observations. They helped me to see myself and analyze my teaching strategies.

In her interview, the principal echoed many of the comments of the teachers. She said that it was important to stay focused as a school on literacy and that a major challenge was developing the reading comprehension of ELL students. She mentioned that guided reading, thinking maps, and writing in response to reading were consistent approaches being used at the school to focus on comprehension. She reported that the school was working on using the analysis of multiple assessments and that data were important to use to inform classroom instructional planning. She valued the collaboration developed through the study-group process, the collaborative dialogue about reading, the discussion about assessment and lesson planning, and the importance of using research. Howard was fortunate to have a principal who was informed about the teachers’ classroom instructional practices and professional growth through study-group activities. Also, she shared the same goals as the teachers.

Teachers consistently mentioned the valuable assistance they received from the CIERA literacy coordinator, dubbing her their “CIERA guru.” Teachers appreciated that she was well informed about the latest research, was organized, and was enthusiastic. They also valued the other members of the leadership team, commenting, “I know there is always help out there to listen. If you had a question, someone will gladly come up to model.” Finally, teachers saw teacher leadership as a shared responsibility, commenting that “Teachers also take on leadership by talking to each other.”

In summary, the teachers at Howard were focusing on improved reading comprehension as a schoolwide goal. In study groups they were learning new approaches to teach reading comprehension. They were reflecting on their reading comprehension instruction as part of the study-group process. They were honing their observation skills as an assessment tool to better understand students’ growth and needs, especially as related to higher level talk, writing about text, and use of comprehension strategies. They were using the observation data as one more tool to give them feedback on their teaching. The teachers, teacher leaders, and principal at Howard were working together as a collaborative learning community, and they were observing the growth

they had hoped to see in their students' reading comprehension.

Overall discussion

Summary of findings

One major purpose of this study was to determine classroom- and school-level variables that accounted for students' growth in reading and writing in schools attempting to improve their reading instruction. We found several. Looking at classroom-level variables across a single year, we found that higher level questioning was positively related to students' reading growth whereas a high level of rote comprehension skill practice was negatively related. Coaching was positively related to students' writing growth. It is important to note that both high-level questioning and coaching were two of the evidence-based classroom practices emphasized in the CIERA School Change Framework. An examination of classroom practices across two years revealed that in schools implementing the reform effort reasonably well, teachers were changing their teaching in the direction of more high-level questioning and more coaching; by contrast, teachers in schools that were not very successful in implementing the framework exhibited few changes in their classroom practices. These findings stress the importance of evidence-based, reflective professional development. Teachers in the high-reform-effort schools appeared to be more attuned to the research on effective reading instruction, and they used this information, along with data on their teaching and their study-group activities, to implement more effective teaching practices.

A second purpose of this study was to determine if the CIERA School Change Framework, as a school improvement tool, had an impact on students' reading and writing growth. At the school level, reform effort was positively related to students' reading growth. The more elements of the CIERA School Change Framework a school implemented, the greater the growth in students' reading comprehension and fluency. This effect was small but significant when examined across a single year; however, this effect was moderately large (65% of the between school variance for reading comprehension, $ES = .49$) when examined across a two-year period.

This difference in impact across one versus two years highlights the importance of sustained school improvement efforts. Substantive change in these schools took root gradually, not suddenly.

Growth in students' reading scores as well as change in classroom teaching practices came in small increments from one year to the next. There were no quick fixes and no magic bullets in these schools—only hard work, persistence, and professional commitment.

The reform-effort findings also add to our knowledge about the complexity of effective school-wide reading improvement. Schools are not going to succeed at a reform effort just because they decide to "try something new." Unfortunately, about a third of the schools in this project were not very successful in implementing the components of the CIERA School Change Framework even though teachers had voted to engage in the reform effort. While disappointing, this finding is not surprising; a similar percentage was reported by Bryk et al. (1998) in their study of school reform in Chicago. Generally, the low-reform schools were not accomplishing the goals they had set out to implement because they lacked commitment to the process as well as perseverance. These schools tended to lack principal support, and usually no teacher leader emerged to keep the reform effort moving forward.

Despite the best of intentions, schools often get sidetracked in one of two ways—either they search for a single, magic bullet to solve all their ills, or, in response to a multitude of pressures from state or district mandates, they shift from topic to topic, hoping that somehow their actions will add up to a solution. A school faculty may decide to increase time for reading instruction, *or* put a new delivery model in place, *or* purchase a new reading curriculum, *or* focus on getting students ready for the state-mandated reading test. At the classroom level, a teacher may implement a new but idiosyncratic teaching routine, *or* switch students with a colleague to meet students at their instructional level in more homogeneous groupings, *or* send the students with the greatest instructional needs to a resource teacher who can apply the latest intervention, *or* put extra effort into getting parents more involved.

There is nothing inherently misguided about any of these individual efforts. Too often, however, they serve as a substitute for what really matters—sustained collaborative work with colleagues in schoolwide efforts and refocused classroom instruction to maximize students' cognitive engagement in their literacy learning. At school level, teachers and principals need to improve shared leadership *and* collaborate in the delivery of reading instruction with a model that puts the students first *and* use data to inform instruction and improve teaching practices, *and* they need to engage in collaborative, job-

embedded professional development. Within classrooms, teachers must reflect on their instruction *and* implement research-based practices that focus on providing support to promote active student involvement in high-level cognitive tasks *and* develop and maintain high expectations for student learning. To achieve all of these goals simultaneously requires hard work, deep analyses of teaching and learning, and a commitment to collective problem solving; teachers who accept the challenge of this agenda will need to change the way they work with their colleagues and their students.

However, with commitment to a sustained school improvement effort, teachers and principals can realize the growth in students' reading ability that they are under so much pressure to achieve. In this project, the success in implementing the reform effort within a school made a significant contribution to growth in student achievement.

It is encouraging to note that in an organization as complex as a school faculty, individuals were able to come together as a community to use data on their teaching practices and participate in focused study-group activities to improve reading instruction. The higher reform-effort schools also typically had one strong and respected teacher leader who was persistent in helping teachers examine the data linking students' growth to classroom practices. Typically, this leader also steered teachers into study-group topics that were likely to make a difference. In most of the high-reform-effort schools, the teacher leader received support and assistance from a group of teachers serving as a teacher leadership team and from the school principal, even if and when the principal was not actively involved in the professional development.

The importance of school community and the efficacy of internal and external facilitation in this work suggest that there are important roles for many players in facilitating school reform. Clearly teacher leaders and principals matter greatly in the process. But district administrators can help by giving schools the degrees of freedom needed to really focus on a single professional development goal. University faculty might assist in two ways: (a) by supporting professional development with new techniques or strategies and (b) by preparing new teachers to work effectively with peers so that they take a collaborative disposition into their work as teachers. Policymakers can also play a role by insisting on policy levers (e.g., systems of standards, assessments, and accountability schemes) that focus on ends and outcomes and leave issues of means (how the ends are achieved) to local schools.

Limitations

This study was limited to 13 schools that were either in their first or second year of the reform effort. Although one may have expected the second-year schools to show more reading growth within a year than first-year schools, this did not prove to be the case.

A second limitation is that resource constraints restricted this study of school improvement to two years. High-reform-effort schools were beginning to see positive changes in students' reading growth, but they had only started on a long journey. Effective school improvement is a complex, multiyear process (Fullan, 1999), and we must be willing to stay the course in these schools for at least five or six years in order to fully understand the nature of change.

Although teachers were randomly selected to participate in the collection of observation and interview data, they had the option (as do all teachers given current guidelines for the protection of human subjects) to decline to participate. Thus, a limitation of this study is that only randomly selected teachers who agreed to participate were part of the teacher data collection.

Because this was a large-scale study including 92 teachers in 13 schools, data on teaching practices were limited to observations of three 1-hour reading lessons per teacher. It would have been preferable to look at the entire literacy block on three occasions or to increase the number of days of observation, but neither option was possible, again due to resource constraints. Even so, the fact that teacher variables explained a substantial amount of the between classroom variance suggests that even this modest sampling approach was genuinely sensitive to consequential variations in instruction and supports the conclusion that teacher practices matter greatly.

Because it would have been a burden to teachers and parents to obtain consent on all students and resource constraints limited the number of students we could assess, we were unable to test all students to determine reading ability (e.g., high, average, or low). Instead we had teachers use their judgment to break their students into the top, middle, and lowest thirds in terms of reading ability, and from those thirds we randomly selected students to participate in the study. Even though teacher judgment has long been acknowledged as a good predictor of achievement test scores, its use should be acknowledged as another limitation of the study.

Questions for further research

Because some schools in the project were not very successful in implementing the components of

the CIERA School Change Framework, we ended our work vexed by the question of how schools can be helped when there is teacher buy-in but no real leadership from a principal or no substantial teacher leadership to keep a reading reform effort moving forward. Many U.S. schools will not have the strong, democratic leadership (Bryk et al., 1998) that is seen as necessary for a school to succeed in implementing a significant reform effort. Giving up on these schools, however, is not an acceptable option. Perhaps district leaders should bear responsibility for seeing that effective school leadership is fostered within schools; or perhaps universities need to rethink their leadership preparation programs to ensure that the topic of leadership for curriculum and pedagogy is a prominent programmatic goal. Clearly, we need more research that examines the appropriate curricular leadership roles for principals and teacher leaders.

The successful schools in this project saw steady, but not dramatic, changes in their teaching practices and student achievement across two years. Clearly, even in these improving schools more time was needed to become a successful school. This raises two research-related questions: (a) How can a school be encouraged or motivated to stay in a reform effort over the long haul? (b) How can the educational research enterprise generate the resources needed to carry out long-term (5–10 years) studies of school change? This is an important policy as well as a research question. So often a school tries a new approach for one or two years and then moves on to something new. Again, it seems that district leadership may have a role to play in stabilizing the leadership within schools and licensing schools, principals, and teacher leaders to focus their energies on systemic issues of curriculum and pedagogy rather than quick fixes, such as teaching to the test, designed to achieve short-term gains in test scores.

Conclusions

The CIERA School Change Framework provided structure to and support for school improvement in reading. The framework appeared to help teachers attend simultaneously to the complexity of both school-level and classroom-level efforts. Nevertheless, it was teachers' collaboration, teachers' decisions about what to study, and teachers' perseverance that made the difference in the more successful schools. This approach to school improvement will not succeed in all schools, especially if teacher buy-in or leadership is lacking. Nonetheless, the study does demonstrate that a framework for change that fosters "homegrown" reading improve-

ment efforts can be used successfully by schools to promote achievement growth. When teachers collaborate, engage in ongoing, reflective professional development, and use data to improve teaching practice, they can achieve significant growth in their students' reading achievement. We close by noting the irony of providing evidence that homegrown models of reform in the United States can work at a time when some state and federal policies seem to be focused on dramatically limiting the choices schools have at their disposal in meeting legislative mandates to improve reading achievement.

BARBARA M. TAYLOR is Guy Bond Professor of Reading in the Department of Curriculum and Instruction at the University of Minnesota, where she is currently heading up the Minnesota Reading First Professional Development Program. Her research interests include school reform in reading, effective professional development in reading, early reading intervention, and reading comprehension instruction. She can be contacted at 338 Peik Hall, University of Minnesota, 159 Pillsbury Drive SE, Minneapolis, MN 55455, USA, or by e-mail at bmtaylor@umn.edu.

P. DAVID PEARSON is a professor of Language and Literacy Education in the Graduate School of Education at the University of California, Berkeley, where he also serves as dean. His research interests include reading curriculum, instruction, and assessment, especially in low-income schools. He can be contacted at 1501 Tolman Hall, GSE, UC Berkeley, Berkeley, CA 94720-1670, USA, or by e-mail at ppearson@socrates.berkeley.edu.

DEBRA S. PETERSON is the assistant project director for the Minnesota REA/Reading First Professional Development Program. She has worked with elementary schools to improve classroom reading instruction and implement schoolwide reform through the CIERA School Change project, the Minnesota Reading Excellence Act, and the Reading First project. She is also a teaching specialist in the Department of Curriculum and Instruction at the University of Minnesota and can be contacted at 245 Peik Hall, 159 Pillsbury Drive SE, University of Minnesota, Minneapolis, MN 55455, USA, or by e-mail at debra.s.peterson-2@umn.edu.

MICHAEL C. RODRIGUEZ is an assistant professor in the Department of Educational Psychology at the University of Minnesota, where he teaches courses in measurement, statistics, and evaluation. His current research interests include classroom assessment, item writing and item format effects, and reliability theory. He can be contacted at 206 Burton Hall, 178 Pillsbury Drive SE, University of Minnesota, Minneapolis, MN 55455, USA, or by e-mail at mcrdz@umn.edu.

REFERENCES

- ALLINGTON, R.L., & JOHNSTON, P.H. (2002). *Reading to learn: Lessons from exemplary fourth-grade classrooms*. New York: Guilford.
- BERGMAN, J.L. (1992). SAIL—A way to success and independence for low-achieving readers. *The Reading Teacher*, 45, 598–602.
- BLACHOWICZ, C., & OGLE, D. (2001). *Reading comprehension: Strategies for independent learners*. New York: Guilford.
- BROPHY, J. (1973). Stability of teacher effectiveness. *American Educational Research Journal*, 10, 245–252.
- BROWN, R., PRESSLEY, M., VAN METER, P., & SCHUDER, T. (1996). A quasi-experimental validation of transactional strategies

- instruction with low achieving second grade readers. *Journal of Educational Psychology*, 88, 18–37.
- BRYK, A.S., & RAUDENBUSH, S.W. (1992). *Hierarchical linear models*. Newbury Park, NJ: Sage.
- BRYK, A.S., SEBRING, P.B., KERBOW, D., ROLLO, S., & EASTON, J.Q. (1998). *Charting Chicago school reform: Democratic localism as a level for change*. Boulder, CO: Westview.
- CALFEE, R. (2003, May). *Reading performance in California's schools*. Paper presented at the Conference of the University of California Literacy Consortium, Berkeley.
- CERVETTI, G.N., JAYNES, C., & PEARSON, P.D. (2002, December). *Fluency: Its nature and impact*. Paper presented at the National Reading Conference, Miami, FL.
- CHARLES A. DANA CENTER, UNIVERSITY OF TEXAS AT AUSTIN. (1999). *Hope for urban education: A study of nine high-performing, high-poverty urban elementary schools*. Washington, DC: U.S. Department of Education, Planning and Evaluation Service.
- DENO, S. (1985). Curriculum-based measurement: The emerging alternative. *Exceptional Children*, 52, 219–232.
- DESIGNS FOR CHANGE. (1998). *Practices of schools with substantially improved reading achievement*. Retrieved June 16, 2004, from <http://www.designsforchange.org>
- DUFFY, G.G., ROEHLER, L.R., SIVAN, E., RACKLIFFE, G., BOOK, C., MELOTH, M.S., ET AL. (1987). Effects of explaining the reasoning associated with using reading strategies. *Reading Research Quarterly*, 20, 347–368.
- DUNKIN, M., & BIDDLE, B. (1974). *The study of teaching*. New York: Holt, Rinehart & Winston.
- FLANDERS, N. (1970). *Analyzing teacher behavior*. Reading, MA: Addison-Wesley.
- FLOOD, J., LAPP, D., SQUIRE, J.R., & JENSEN, J.M. (Eds.). (2003). *Handbook of research on teaching the English language arts* (2nd ed.). Mahwah, NJ: Erlbaum.
- FULLAN, M. (1999). *Change forces: The sequel*. Philadelphia: Falmer.
- FULLAN, M. (2002). Educational reform as continuous improvement. In W.D. Hawley & D.L. Rolie (Eds.), *The keys to effective schools: Educational reform as continuous improvement* (pp. 1–9). Thousand Oaks, CA: Corwin.
- FULLAN, M., & HARGREAVES, A. (1996). *What's worth fighting for in your school*. New York: Teachers College Press.
- GREENWOOD, C.R., CARTA, J.J., KAMPS, D., & DELQUADRI, J. (1995). *Ecobehavioral assessment systems software (EBASS) practitioner's manual* (Version 3.0). Kansas City: Juniper Garden's Children's Project, University of Kansas.
- HAWLEY, W.D. (Ed.). (2002). *The keys to effective schools: Educational reform as continuous improvement*. Thousand Oaks, CA: Corwin.
- HOFFMAN, J.V. (1991). Teacher and school effects in learning to read. In R. Barr, M.L. Kamil, P.B. Mosenthal, & P.D. Pearson (Eds.), *Handbook of reading research* (Vol. 2, pp. 911–950). New York: Longman.
- KAMIL, M.L., MOSENTHAL, P.B., PEARSON, P.D., & BARR, R. (Eds.). (2000). *Handbook of reading research* (Vol. 3). Mahwah, NJ: Erlbaum.
- KILLION, J. (2002). *What works in the elementary school: Results-based staff development*. Oxford, OH: National Staff Development Council.
- KNAPP, M.S. (Ed.). (1995). *Teaching for meaning in high-poverty classrooms*. New York: Teachers College Press.
- KREFT, I., & DELEEUW, J. (1998). *Introducing multilevel modeling*. London: Sage.
- LANGER, J.A. (1999). *Beating the odds: Teaching middle and high school students to read and write well* (Rep. No. 12014). Albany, NY: National Council of Teachers of English.
- LANGER, J.A. (2000). Excellence in English in middle and high school: How teachers' professional lives support student achievement. *American Educational Research Journal*, 37, 397–439.
- LEE, V.E., & BRYK, A. (1989). A multilevel model of the social distribution of educational achievement. *Sociology of Education*, 62, 172–192.
- LEE, V.E., & SMITH, J.B. (1997). High school size: Which works best and for whom? *Educational Evaluation and Policy Analysis*, 19, 205–227.
- LEIN, L., JOHNSON, J.F., & RAGLAND, M. (1997). *Successful Texas schoolwide programs: Research study results*. Austin: Charles A. Dana Center, University of Texas at Austin.
- LIEBERMAN, A., & MILLER, L. (2002). Transforming professional development: Understanding and organizing learning communities. In W.D. Hawley (Ed.), *The keys to effective schools: Educational reform as continuous improvement* (pp. 74–85). Thousand Oaks, CA: Corwin.
- LOUIS, K.S., & KRUSE, S. (1995). *Professionalism and community in schools*. Thousand Oaks, CA: Corwin.
- MAYER, D.P. (1998). Do new teaching standards undermine performance on old tests? *Educational Evaluation and Policy Analysis*, 20, 53–73.
- NATIONAL INSTITUTE OF CHILD HEALTH AND HUMAN DEVELOPMENT. (2000). *Report of the National Reading Panel: Teaching children to read: An evidence-based assessment of the scientific research literature on reading and its implications for reading instruction* (NIH Publication No. 00-4769). Washington, DC: U.S. Government Printing Office.
- NEUMAN, S.B., & DICKINSON, D. (Eds.). (2001). *Handbook of early literacy research*. New York: Guilford.
- NEWMANN, F.M. (2002). Achieving high-level outcomes for all students: The meaning of staff-shared understanding and commitment. In W.D. Hawley (Ed.), *The keys to effective schools: Educational reform as continuous improvement* (pp. 28–42). Thousand Oaks, CA: Corwin.
- PRESSLEY, M. (2002). *Reading instruction that works: The case for balanced teaching* (2nd ed.). New York: Guilford.
- PRESSLEY, M., WHARTON-MCDONALD, R., ALLINGTON, R.L., BLOCK, C.C., MORROW, L., TRACEY, D., ET AL. (2001). A study of effective first-grade literacy instruction. *Scientific Studies of Reading*, 5, 35–58.
- PUMA, M.J., KARWEIT, N., PRICE, C., RICCIUITI, A., THOMPSON, W., & VADEN-KIERNAN, M. (1997). *Prospects: Final report on student outcomes*. Washington, DC: U.S. Department of Education, Planning and Evaluation Services.
- RAUDENBUSH, S.W., & BRYK, A.S. (2002). *Hierarchical linear models: Applications and data analysis methods* (2nd ed.). Thousand Oaks, CA: Sage.
- RAUDENBUSH, S.W., BRYK, A.S., CHEONG, Y.F., & CONGDON, R. (2000). *HLM: Hierarchical linear and nonlinear modeling* (Version 5) [Computer software]. Chicago, IL: Scientific Software International.
- ROSENTHAL, R., & ROSNOW, R.L. (1984). *Essentials of behavioral research: Methods and data analysis*. New York: McGraw-Hill.
- SCANLON, D.M., & GELZHEISER, L.M. (1992). *Classroom observation manual*. Unpublished manuscript, University at Albany, State University of New York, NY.
- SIMPSON, R.J., RAUDENBUSH, S.W., & EARLS, F. (1997). Neighborhoods and violent crime: A multilevel study of collective efficacy. *Science*, 277, 918–924.
- SNOW, C., BURNS, S., & GRIFFIN, P. (Eds.). (1998). *Preventing reading difficulties in young children*. Washington, DC: National Academy Press.
- STALLINGS, J., & KASKOWITZ, D. (1974). *Follow through classroom observation evaluation 1972–73* (SRI Project URU-7370). Stanford, CA: Stanford Research Institute.
- TAYLOR, B.M., & PEARSON, P.D. (2000). *The CIERA school change classroom observation scheme*. Minneapolis: University of Minnesota.
- TAYLOR, B.M., & PEARSON, P.D. (in press). Using study groups and reading assessment data to improve reading instruction within a school. In S. Paris & S. Stahl (Eds.), *Children's reading comprehension and assessment*. Mahwah, NJ: Erlbaum.
- TAYLOR, B.M., PEARSON, P.D., CLARK, K., & WALPOLE, S. (2000). Effective schools and accomplished teachers: Lessons about primary grade reading instruction in low-income schools. *Elementary School Journal*, 101, 121–166.
- TAYLOR, B.M., PEARSON, P.D., PETERSON, D.S., & RODRIGUEZ, M.C. (2003). Reading growth in high-poverty classrooms: The influence of teacher practices that encourage cognitive engagement in literacy learning. *Elementary School Journal*, 104, 3–28.
- TAYLOR, B.M., PETERSON, D.P., PEARSON, P.D., & RODRIGUEZ, M.C. (2002). Looking inside classrooms: Reflecting on the “how” as well as the “what” in effective reading instruction. *The Reading Teacher*, 56, 270–279.
- TAYLOR, B.M., PRESSLEY, M., & PEARSON, P.D. (2002). *Research-supported characteristics of teachers and schools that promote reading achievement*. Washington, DC: National Education Association.
- VALLI, L., & HAWLEY, W.D. (2002). Designing and implementing school-based professional development. In W.D. Hawley (Ed.), *The keys to effective schools: Educational reform as continuous improvement* (pp. 86–96). Thousand Oaks, CA: Corwin.

VENEZKY, R.L., & WINFIELD, L. (1979). *Schools that succeed beyond expectations in teaching reading* (Tech. Rep. No. 1). Newark: Department of Educational Studies, University of Delaware.

WENGLINSKY, H. (1998). Finance equalization and within-school equity: The relationship between education spending and the social distribution of achievement. *Educational Evaluation and Policy Analysis*, 20, 269–283.

WOLF, S.A., BORKO, H., ELLIOTT, R.L., & MCIVER, C.M. (2000). "That dog won't hunt!" Exemplary school change efforts within the Kentucky reform. *American Educational Research Journal*, 37, 349–393.

YSSELDYKE, J., & CHRISTENSON, S. (1993–1996). *TIES: The instructional environment system—II*. Longmont, CO: Sopris West.

AUTHORS' NOTE

This research was conducted as part of CIERA, the Center for the Improvement of Early Reading Achievement, and supported under the Educational Research and Development Centers Program, PR/Award No. R305R70004, as administered by the Office of Educational Research and Improvement, U.S. Department of Education. However, the contents of the article do not necessarily represent the positions or policies of the National Institute on Student Achievement, Curriculum, and Assessment; the National Institute on Early Childhood Development; or the U.S. Department of Education, and readers should not assume endorsement by the federal government.

HLM RESULTS AND TABLES

Interpreting HLM results and tables

There are three pieces of HLM output that are relevant to readers. These include variance components for the initial random effects and final random effects and final fixed-effects coefficients. The initial random-effects variance components are based on the degree to which random effects vary prior to including predictors (which potentially explain variation). Random effects include the classroom means and school means, which are allowed to randomly vary in these models. Because of this partitioning of variance, the percentage of variation between classrooms and schools can be computed. Once predictors are entered into the model, these random-effects variance components are potentially reduced if the predictors actually explain variation in the random effects—these are the final random-effects variance components. Based on these figures, the percentage of variance explained can be computed (percentage of reduction in variance from initial to final models).

The last section of each output table includes the final fixed-effects coefficients. These are equivalent to regression coefficients (regression weights), based on the combined effects of each predictor, as they explain variation in student residuals, classroom means, or school means. These coefficients have associated standard errors from which t tests can be computed with known p values for hypothesis testing. Which source of variation each predictor explains is noted in parentheses in the HLM tables.

Effect size computation

Unless otherwise noted, the effect size was calculated by dividing the coefficient of the predictor variable in the final model by the standard deviation (square root of the variance component) of the appropriate classroom, student, or school mean in the base model. For example, the effect size for reform effort accounting for spring reading comprehension scores was calculated as $1.34/4.55 = 0.29$, where 1.34 is the coefficient for reform effort in the final model and 4.55 is the square root of 20.68, the variance for schools' means in the base model. This procedure has been used in other studies using HLM where effects are reported in SD units for the variation at the relevant level (see Lee & Smith, 1997, for a discussion).

In some cases, observation scores were a ratio, ranging from .00 to 1.00. We calculated the effect size for a significant classroom observation variable in terms of 1 standard deviation of change in that classroom observation variable. This approach to calculating effect size was used for all significant classroom observation variables. For example, the effect size for comprehension skills accounting for spring reading comprehension scores was calculated as $(13.78/7.05)/.14 = 0.27$, where 7.05 is the square root of 49.72, the variance for classroom means in the base model, and .14 is the standard deviation for the observed incidence of comprehension skills (see Tables 7 and 8).